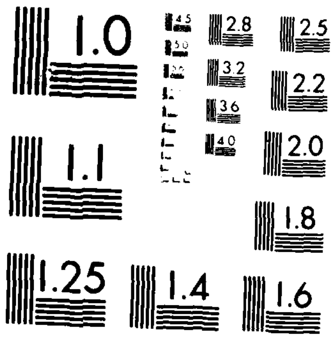
AD-A152 127 THE DEVELOPMENT OF AN ENLISTMENT STANDARDS MODEL FOR THE NAVY AVIATION MACHINIST'S MATE (AD) RATING(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D A OSLUND ET AL. 1/2

UNCLASSIFIED JUN 84 F/G 5/9 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

(2)

AD-A152 127

NAVAL POSTGRADUATE SCHOOL

Monterey, California

DTIC
ELECTE
APR 5 1985
S D
B

THESIS

THE DEVELOPMENT OF AN
ENLISTMENT STANDARDS MODEL
FOR THE NAVY
AVIATION MACHINIST'S MATE (AD) RATING

by

Dwayne A. Oslund

and

J. S. A. Clark

June 1984

Thesis Advisor: W. E. McGarvey

DTIC FILE COPY

Approved for public release, distribution unlimited.

85 03 18 016

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A152 127	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Development of an Enlistment Standards Model for the Navy Aviation Machinist's Mate (AD) Rating		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis June 1984
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dwayne A. Oslund and J. S. A. Clark		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		12. REPORT DATE June 1984
		13. NUMBER OF PAGES 113
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Enlistment Standards, Enlistment Models, Regression Analysis, Discriminant Analysis, Utility Analysis, Model Validation, Selection Standards, Personnel Selection

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this study is to present analytic techniques for developing enlistment standards models which attempt to improve upon existing methods. Alternative criteria for measuring successful operational performance, and a means of measuring utility are also provided. Another purpose of this study is to discover if the Navy's system of selecting personnel for the Aviation Machinist's Mate (AD) rating may be improved.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE Unclassified

S N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Two criteria were utilized in developing these models--length of service, and a composite measure of success that considers length of service, highest paygrade achieved, and reenlistment eligibility. Measures on individual's at the time of enlistment are used as predictor and discriminating variables. Six models are developed and analyzed using regression and discriminant techniques. Utility analysis is conducted on each of these models as a means for measuring their usefulness in monetary terms. Recommendations for future research are also presented.

Approved for public release; distribution unlimited.

The Development of an Enlistment Standards Model for the Navy Aviation Machinist's Mate (AD) Rating

by

Dwayne A. Oslund
Lieutenant Commander, United States Navy
B.S., Oregon State University, 1971

and

Captain J. S. A. Clark, RAE
Australian Army
B.E. (Civil), University of New South Wales, 1976

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
June 1984

Authors: Dwayne A. Oslund

J. Clark

Approved by: William E. Whitney
Thesis Advisor

Richard A. Elster
Second Reader

Richard A. Elster
Chairman, Department of Administrative Sciences

Kneale T. Marshall
Dean of Information and Policy Sciences

ABSTRACT

The purpose of this study is to present analytic techniques for developing enlistment standards models which attempt to improve upon existing methods. Alternative criteria for measuring successful operational performance, and a means of measuring utility are also provided. Another purpose of this study is to discover if the Navy's system of selecting personnel for the Aviation Machinist's Mate (AD) rating may be improved.

Two criteria were utilized in developing these models--length of service, and a composite measure of success that considers length of service, highest paygrade achieved, and reenlistment eligibility. Measures on individual's at the time of enlistment are used as predictor and discriminating variables. Six models are developed and analyzed using regression and discriminant techniques. Utility analysis is conducted on each of these models as a means for measuring their usefulness in monetary terms. Recommendations for future research are also presented.

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

For the remainder of this decade and beyond, the Navy is faced with the difficult problem of attracting and retaining sufficient personnel to meet its ever increasing manpower requirements. The fleet is expanding toward 600 ships while the available manpower pool of 17 to 21 year old men and women is projected to decline. Each year, millions of dollars are spent recruiting, training and maintaining enlisted personnel. Numerous enlistment standards models have been developed to improve the screening, selection and assignment processes for all Navy ratings. Continuing enlistment standards research is important since it may improve manpower planning, reduce attrition, enhance job performance, and lengthen careers. It is through such research that the ultimate goal of increased readiness at lower cost may be realized.

A. PURPOSE OF ANALYSIS

This study attempts firstly to improve upon the methodology presently utilized to develop enlistment standards models. In particular, different techniques for developing such models are presented, along with alternative criteria for measuring successful performance. A means of measuring the utility, or usefulness, of such efforts is also provided. An attempt has been made to present these methods in a clear manner so that those researchers who follow may more readily understand the process. The analysis expands upon the experience of numerous similar efforts, including several graduate theses prepared at the Naval Postgraduate School and many research projects conducted under the

auspices of the Navy Personnel Research and Development Center (NPRDC) and the Center for Naval Analyses (CNA).

The secondary purpose of this study is to discover if the selection standards for one particular Navy rating may be improved upon by analyzing data available at the time of enlistment. Most predictive models developed to date have focused on successful completion of technical training schools, or on attrition. This study is aligned with a more recent analytic trend of attempting to predict successful operational performance in the fleet. This approach is considered most appropriate in today's highly technical Navy. The tremendous cost, in terms of both dollars and time, associated with training and retaining Navy personnel demands maximum return on investment. By focusing on operational success to develop a larger, more experienced career force, there exists the potential to reduce the burden of recruiting and training new enlistees.

E. RATING SELECTED FOR ANALYSIS

To accomplish the above stated purposes, data pertaining to operational performance of personnel in the Aviation Machinist's Mate (AD) rating were analyzed. ADs are aircraft engine mechanics who inspect, adjust, test, repair and overhaul engines used in all Navy airplanes and helicopters. ADs also perform routine maintenance, prepare aircraft for flight, and assist in handling aircraft on the ground or aboard ships. They perform maintenance and servicing on either jet or reciprocating engines, and on subsystems such as fuel, oil, induction, compression, combustion, turbine and exhaust. Other AD functions include supervising maintenance, analyzing fuel and oil samples, keeping records, evaluating engine performance, and maintaining accessory components, drive systems and gear boxes.

Aviation Machinist's Mates are assigned primarily to Naval Aviation squadrons or to Aircraft Intermediate Maintenance Departments. These assignments may be either afloat or ashore. ADs may also be assigned as instructors in training activities, and they are eligible to volunteer for flight duty as aircrewmen. [Ref. 1]

Presently, there are over 13,000 men and women assigned to the AD rating. Since ADs are assigned to virtually every Navy aviation unit, they represent a vital element in ensuring a high degree of aircraft readiness is maintained. As such, the overall mission effectiveness of Naval Aviation units is directly linked to the quality and performance of members of the AD rating.

C. ORGANIZATION OF THIS STUDY

This chapter has discussed the purpose of this study, and described the AD rating and its importance to the Navy. The next chapter will provide background information on enlistment standards research, and present in general terms the evolution of predictor and criterion variables that emerged during the development of the models contained in this research. Chapter III describes the data base and AD data set that provided specific measures of operational performance for analysis and model formulation. Chapter IV presents the three statistical analysis techniques employed in developing six enlistment standards models. Chapter V provides the method and results of the utility analysis conducted on these models. Utility analysis represents a means by which the usefulness of similar efforts may be measured. Chapter VI draws conclusions from the analysis and presents recommendations for further research.

II. SELECTION OF VARIABLES

This chapter gives a brief description of some of the selection procedures in use at the time of the data collection. The results of previous research on predicting job performance are reviewed and predictor and criterion variables that have been shown to be useful are identified.

A. SELECTION BACKGROUND

At the time the data used in this analysis was collected, the Navy considered a number of applicant characteristics to guide selection and classification decisions. These characteristics included education, high school degree status, age, number of dependents, scores on the 12 Armed Services Vocational Aptitude Battery (ASVAB) subtests, and some composite scores. The Armed Forces Qualification Test (AFQT) is one of these composite scores, and an applicant's score on the AFQT depended on the sum of his scores on the ASVAB subtests Arithmetic Reasoning (AR), Spatial Perception (SP), and Word Knowledge (WK). The AFQT score was reported as a percentile--a score of 80 meant that an applicant's total score on the three subtests was higher than the scores achieved by 79 percent of his peers. The AFQT percentile score was also used to classify the applicant into one of five mental categories or AFQT groups. For example, those with a score of 90 or better were classified in mental group I, and those with a score of 10 or less were classified as group Vs.

Another composite score is the Success Chances of Recruits Entering the Navy (SCREEN) score. This score is derived from the personal characteristics of age at entry,

years of schooling, whether or not the applicant had dependents, and AFQT percentile score. This composite score has been used by recruiters since October 1976 in determining an applicant's eligibility to enlist. [Ref. 2]

A final composite score that is used in classifying recruits to the AD rating is made up of the sum of the recruit's standardized scores on the ASVAB subtests Arithmetic Reasoning (AR), Electronic Information (EI), General Science (GS), and Mathematical Knowledge (MK). A minimum score of 190 on this composite was necessary for a recruit to enter the AD rating.

E. REVIEW OF PREVIOUS MILITARY STUDIES

Studies on predicting military job performance have mainly concentrated on the survivability of recruits through various stages of their military careers. These stages include recruit training, Class "A" School, first two years of enlistment and first term of enlistment.

Lurie used number of dependents, years of education and AFQT score to predict the performance of the Electronic's Technician (ETN) and Ship's Serviceman (SH) ratings. He found that for the SH rating, non-high school graduates with lower AFQT scores were promoted faster than those with higher scores, although AFQT score had no impact on first term survival. The AFQT score did aid in predicting advancement and survival for members of the ETN rating. [Ref. 3] In another study of eight year survival rates, Lurie found that education level was the most important predictor. Interestingly he also found that mental group I recruits had the worst record in surviving Class "A" School. [Ref. 4]

A study on the factors affecting first term survival and retention behavior of Machinist's Mates (MM) and Boiler

Technicians (BT) was conducted by Fletcher in 1979. He found that BTs with greater than 11 years of education had greatly improved chances of surviving their first term of enlistment. For MMs, those in the lowest and highest mental groups had greater survival probability than others. For both ratings, older men had a higher probability of survival. In relation to reenlistment, those BTs with 12 or more years of education had a low probability of reenlistment, while with MMs, having a dependent increased the probability of reenlistment. [Ref. 5]

Studies of military job performance have also investigated the effect of the Delayed Entry Program (DEP) on survival. Lockman found that if recruit quality (as measured by SCREEN) and training guarantees were controlled for, those who were in the DEP for three or more months had the highest survival rates [Ref. 6]. Thomason found that DEP, age, education, recruit training location, race, number of dependents, mental group and follow on tour assignments had varying degrees of significance in determining first term survivability [Ref. 7].

More recent studies have favored the use of multiple, rather than single measures of job performance. This is because it is rare that a single measure adequately covers the full scope of job performance. One approach has been to expand the survivability criteria to include other measures of job performance, such as eligibility to reenlist and the achievement of certain paygrades. A continuous criterion is not available under this approach; sailors are either categorized as a success or as a failure. Nesbitt [Ref. 8] and Snyder and Bergazzi [Ref. 9] defined various degrees of success or failure in their studies in an effort to generate more variability on the criterion.

C. CRITERION AND PREDICTOR VARIABLES

In most cases when a single job performance measure (criterion) has been used in previous research, a measure of survival has been the overwhelming choice. This is because such a criterion is readily measurable, is continuous, and is of importance to the Navy since the costs associated with attrition and subsequent replacement are considerable. Other single criteria have been length of service (LOS), time to promotion, highest rank or grade achieved, retention (as measured by reenlistment choice), and performance at Class "A" Schools.

The common predictors of job performance are education, number of dependents, age, sex, race, ASVAB subtest scores, AFQT scores, mental group, DEP status, and some "after accession" variables such as recruit training location, subsequent dependent status, and initial and follow on duty assignments.

In this study two criteria will be considered. The first will be an LOS criterion and the second will be a composite criterion where success will be defined as completing the first term of enlistment, being eligible for reenlistment, and achieving the paygrade E-4. The candidate predictor variables will be age at entry, sex, race, entry paygrade, education, dependent status, term of enlistment, ASVAB subtest scores, AFQT scores and the composite score to qualify for the AD rating. The specific variables from the AD data set used for analysis, as well as the evolution of the data set, are discussed in the next chapter.

III. DATA BASE DEVELOPMENT

This chapter provides information concerning the master data base and the subset of this master file, the AD data set, that was used in this study. The generation of this AD data set is described in detail, as are the specific predictor and criterion variables discussed in the previous chapter.

A. MASTER FILE

Enlisted history records on over 206,000 non-prior service sailors who entered the Navy during the period 1 September 1976 to 31 December 1978 were compiled by the Defense Manpower Data Center (DMDC) staff. This master file was created by merging data elements from four separate files: (1) the DMDC Cohort file, which is itself a compilation of information from DMDC's Enlisted Master Record and Loss files; (2) a Navy Health Research Center (NHRC) file; (3) a promotion examination file; and (4) a Chief of Naval Education and Training (CNET) file.

The DMDC Cohort file contains personal and demographic data on individuals at the time they entered the service. Additionally, it is updated quarterly by the Military Personnel Commands with active duty or service separation information as appropriate. This file provided over 150 variables to the master file.

The NHRC file contains information on each enlisted member of the Navy who has been or still is on active duty. It is updated quarterly from Navy Military Personnel Command (NMPC) change tape extracts, and follows a service member from date of enlistment to date of discharge. The NHRC file represents approximately 30 variables in the master file.

The promotion examination file includes advancement exam and promotion data, and the CNET file contains information on formal training received by individuals in the data base. Together these files provided over 60 variables to the master data base.

The master file, containing 243 variables, is maintained at the Naval Postgraduate School. The final update to the file includes DMDC data current as of 30 September 1982, and NHRC data current as of July 1982. The program to access data on the selected rating is contained in Appendix A.

B. AD DATA SET

This section describes the evolution of the AD data set that contains the observations and measures analyzed in this study. The AD data set was derived through a number of iterative screens, and new variables were created, in order to remove aberrant observations and to refine the predictor and criterion variables prior to statistical analysis. These iterative steps ultimately reduced the number of cases in the AD data set from 5,562 to 2,820 observations. The programs used to screen the data and to create new predictor and criterion variables are contained in Appendix A. The logic for these processes is discussed in the remainder of this chapter.

1. Screens

Since one purpose of this study was to analyze Aviation Machinist's Mates who had operational experience in the fleet, the first screen performed on the data was to select only those cases whose final DMDC rate (DMDCRATE)[1] appeared in the last master file update as ADs. This means

---

[1]The actual variable name associated with the comment is provided in parentheses throughout the remainder of this chapter.

they were working in the AD rating at either the time of their separation from the service, or at the time of the last file update if they were still on active duty. This screen allowed for people to migrate into the AD rating while ensuring that those cases who left for another rating were excluded from the analysis.

The cases were next screened for ADs with no prior Navy service (PRIORSRV). In addition, individuals who may have changed their social security number (SSNCHNGE) were removed from the sample. These screens ensured that no observations were counted more than once in the analysis.

The observations in the AD data were then screened to select only those people who were tested on ASVAB Forms 5, 6, or 7 (TESTFORM) at enlistment. These test forms were in use during the period in which the individuals in the data set enlisted. Also, those cases whose subtest scores (ASVABs) were impossibly high were eliminated from the data set.

Two other screens were conducted to capture those observations who enlisted in either the Regular Navy or Naval Reserve (SERVACCS), and who were known to have signed up for at least four years <u>active</u> service (RECENLST). It is worth noting that during model development, the term of enlistment measure (TERMENLT) was consistently significant, but with a negative value for the parameter estimate. This indicated that those individuals who enlisted for longer obligated service actually served less time than those who signed up for shorter terms of enlistment. The parameter estimate for term of enlistment changed to a positive value when the RECENLST screen was implemented. Apparently, by screening for four year active duty obligors, the data set then excluded those <u>reservists</u> who were required to serve only three years of their six year obligation on active duty. For these observations, a six year term of enlistment

was an erroneous value for the TERMENLT variable. This important discovery reveals a probable flaw in earlier similar enlistment standards analytic efforts.

Another screen facilitated inclusion of those cases for which there was clear indication of their eligibility to reenlist (ELGREUP1 or ELGREUP3). The final screen in setting up the AD data set included only those separated individuals who could be easily identified by "good" or "bad" interservice separation codes (ISC). Observations with separation codes in the "grey" area (death, hardship discharge, entry into officer programs, or medical disqualification) were removed from the data set since it was felt a legitimate determination of their success or failure could not be made.

Having incorporated these screens, frequency distribution analysis facilitated the removal of aberrant or impossible cases. For example, the maximum length of service between 1 September 1976 and 30 September 1982, the period of the data base, was 72 months. Cases who were listed as having greater than 72 months LOS were removed from the data set.

2. <u>Created Variables</u>

This discussion identifies the variables created in addition to those already in the master data base. Creating these variables facilitated more detailed analysis of observations in the AD data set, and enhanced the enlistment standards model development process. The following comments will also address the reasons for selecting some variables and not others.

a. Predictor Variables

There were several ways that individuals in the master data base might appear in the AD rating during their

career. They may have enlisted in a program to become an AD, taken the AD rating exam, and/or achieved the AD rating through on the job training. To distinguish between the various combinations of these processes, an entry group variable (ENTRYGRP) was created. Frequency analysis of this new variable confirmed that the final DMDC rate of AD screened and selected only those cases who actually ended up as ADs. An effort was made to develop models for various combinations of these entry groups during stepwise regression analysis. However, the derived models were not significant, and they did not improve upon the models ultimately selected for analysis.

A common predictor variable in enlistment standards models is one dealing with education. The measure in the master data base reflecting education level (HYEC) was converted from a qualitative value to a continuous variable (CHYEC) to facilitate statistical analysis. In addition, a dichotomous (0,1) variable was created to reflect attainment of a high school degree (HSDG). During stepwise analysis, which is discussed in the next chapter, each of these two new variables was offered separately as a candidate predictor variable. In nearly every instance, HSDG was shown to be more significant than CHYEC.

Other common predictor variables which measure entry-level attributes are ASVAB subtest scores. To allow the use of these measures of individual characteristics, the scores were standardized, and the created variables (SASVABs) were considered during model development. As mentioned in Chapter II, standardized ASVAB subtest scores are used in various combinations as composite measures. One of these composite variables is AFQT percentile (AFQTPCNT), which also yields AFQT groups (AFQTGRPS). Another composite is the score used to determine eligibility for the AD rating. Two variables were created in the AD data set to

identify this latter composite measure. The first variable created (ADCOMPOS) was a continuous variable which had a value equal to the sum of the four ASVAB standardized subtest scores that make up the composite. The second variable created (ADMINSCR) was a dichotomous variable which distinguished those ADCOMPOS values greater than or equal to 190 from those ADCOMPOS values less than 190. Each time one of these four composite measures was offered as a candidate predictor variable during regression model development, three separate trials were run. One trial contained the composite measure and all 12 SASVAB variables. Another trial contained the composite variable and only those SASVAB variables that did not make up the composite variable. The third trial contained only the composite measure with no SASVAB variables. Additionally, the trials contained either AFQTPCNT or AFQTGRPS, and either ADCOMPOS or ADMINSCR. The purpose of this iterative process was to ensure multicollinearity effects were minimized among the independent variables. During the development of the regression models, AFQTPCNT and ADMINSCR were consistently shown to be more significant than AFQTGRPS and ADCOMPOS respectively. For this reason, they were included among the final candidate predictor variables used in stepwise regression analysis.

Another predictor variable commonly considered by enlistment standards research deals with marital status and dependents. The master file contains a qualitative variable (MRTLDPND) which reflects marital status and number of dependents. This study created a dichotomous variable (DEPENDIS) which distinguishes single individuals from those who are married and/or who have children. Again an iterative process revealed this created variable to consistently be more significant.

The effects of race and sex were also considered in the analysis by creating new variables. The best

variable in the master file to indicate race and ethnic status identified categories of whites, blacks and others (RACE). Since this variable was qualitative, three dummy variables were created (WHITE, BLACK, and OTHER). To allow analysis of the effects of sex, the master file variable (SEX) was converted to a "0,1" variable (NUSEX).

Several other predictor variables were considered and tested for significance and possible inclusion in the final set of candidate predictor variables prior to developing the regression models. Age at enlistment (ENTRYAGE), enlistment paygrade (ENTRPAYG) and term of enlistment (TERMENLT) were among those selected. Many variables were rejected because other measures were better able to capture the desired effects. One particular variable which did not show to be significant was the composite SCREEN variable (SCREEN) discussed in Chapter II. This may be because the components of the SCREEN variable are individually more appropriate for analysis, particularly when the emphasis is on predicting operational performance in the fleet. Similar results were cited by McGarvey [Ref. 10].

The final set of predictor variables created in the AD data set are interaction terms. These variables represent all two-level interactions of the seven variables that met the specified significance level during stepwise regression analysis. The development of these variables is discussed in more detail in Chapter IV.

b. Criterion Variables

As discussed in Chapter II, this study used two criterion variables when developing the six models--length of service measures and success measures. The length of service measure for regression models is a continuous variable (TAFMS1), and for discriminant models is a dichotomous

variable (SUCCTAF).[2] SUCCTAF was assigned a value of 'one' if the value of TAFMS1 was greater than or equal to 48 months, or if the value of TAFMS1 was greater than or equal to 45 months and the individual entered the Navy in October, November or December 1978 (IATEENLT). This was done to ensure those cases who did not have the opportunity to serve 48 months were not improperly classified as failures.

Individuals were considered as successes, for purposes of this analysis, if they served 48 months or longer, achieved paygrade E-4, and were recommended for reenlistment. Again, observations who did not have the opportunity to serve 48 months were also considered successful on the LOS portion of this criterion if they served at least 45 months. The success criterion variable (SUCCESS2) captures these measures by considering SUCCTAF and two other created variables (SUCCPAYG and SUCCREUP).

SUCCPAYG identifies those cases who achieved E-4 as measured by two created variables (PAYGRADE and NUHYPAY). PAYGRADE was created from one of two DMDC variables (PAYGRDE1 or PAYGRDE3) that measure an individual's paygrade at the last file update or upon discharge from the service, as appropriate. NUHYPAY was created by converting an NHRC variable (HYPAYGRD) from a categorical to a numeric variable. Using both DMDC and NHRC measures of paygrade ensured correct classification of an individual on this portion of the criterion.

SUCCREUP, the eligibility to reenlist portion of the success criterion, was derived from the DMDC variable (ELGREUP1) that captured the reenlistment code assigned upon an individual's discharge from the service. Service members on active duty as of the last master file update were considered eligible to reenlist, as long as there was no

----

[2]Discriminant analysis requires the use of categorical vice continuous variables as classification variables.

other information to the contrary. The next chapter will discuss how the information contained in the AD data set was specifically evaluated using three separate statistical analysis techniques.

IV. STATISTICAL ANALYSIS

Three distinct statistical methods were employed in this research: Descriptive Analysis, Regression Analysis and Discriminant Analysis. All methods used Statistical Analysis System (SAS) procedures to analyze the data and develop the models. Table I contains a list of the 46 candidate predictor/discriminating variables used in this study. In all, six sets of variables emerged, and each set was analyzed using both regression and discriminant techniques for comparison. These six sets of predictor/discriminating variables are shown in Table II, along with the appropriate criterion/classification variable. Each method, along with the results, are discussed in the following sections of this chapter. It is worth noting that, while the results may not represent a marked improvement over the selection process in use when individuals in the data set enlisted, the methodology presented may be applied to further analysis of the AD rating or to any other rating in the Navy.

A. DESCRIPTIVE ANALYSIS

Descriptive analysis was accomplished through review of frequency distributions, summary statistics and multivariate correlations.

1. Frequency Analysis

Frequency distributions are summary tables in which data are grouped or arranged into conveniently established numerically ordered classes or categories. The process of data analysis is, therefore, made much more manageable and

TABLE I

Candidate Predictor/Discriminating Variables

Variable	Label
AFQTPCNT	AFQT PERCENTILE (OR EQUIVALENT)
AFQTGRPS	AFQT GROUPS (5,4C,4B,4A,3B,3A,2,1)
ENTRYAGE	AGE OF INDIVIDUAL AT TIME OF ENTRY
ENTRPAYG	ENTRY PAYGRADE (E0--O11)
TERMENLT	TERM OF ENLISTMENT (NO. OF YEARS)
HSDG	HIGH-SCHOOL GRADUATE(1) V. OTHER(0)
DEPENDTS	SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
CNYEC	CONVERTED NUMBER OF YEARS OF EDUCATION
SASVABGI	STANDARDIZED SCORE - GENERAL INFORMATION
SASVABNO	STANDARDIZED SCORE -NUMERICAL OPERATIONS
SASVABAD	STANDARDIZED SCORE - ATTENTION TO DETAIL
SASVABWK	STANDARDIZED SCORE - WORD KNOWLEDGE
SASVABAR	STANDARDIZED SCORE -ARITHMETIC REASONING
SASVABSP	STANDARDIZED SCORE - SPACE PERCEPTION
SASVABMK	STANDARDIZED SCORE - MATH KNOWLEDGE
SASVABEI	STANDARDIZED SCORE - ELECTRONIC INFO
SASVABMC	STANDARDIZED SCORE - MECH COMPREHENSION
SASVABGS	STANDARDIZED SCORE - GENERAL SCIENCE
SASVABSI	STANDARDIZED SCORE - SHOP INFORMATION
SASVABAI	STANDARDIZED SCORE - AUTO INFORMATION
BLACK	(1) BLACK, ELSE (0)
OTHER	(1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX	(1) MALE, (0) FEMALE
ADCOMPCS	AD ASVAB COMPOSITE
ADMINSCR	AD ASVAB COMPOSITE SCREEN
INTER01	DEPENDTS * HSDG
INTER02	DEPENDTS * BLACK
INTER03	DEPENDTS * NUSEX
INTER04	DEPENDTS * TERMENLT
INTER05	DEPENDTS * SASVABAI
INTER06	DEPENDTS * ADMINSCR
INTER07	HSDG * BLACK
INTER08	HSDG * NUSEX
INTER09	HSDG * TERMENLT
INTER10	HSDG * SASVABAI
INTER11	HSDG * ADMINSCR
INTER12	BLACK * NUSEX
INTER13	BLACK * TERMENLT
INTER14	BLACK * SASVABAI
INTER15	BLACK * ADMINSCR
INTER16	NUSEX * TERMENLT
INTER17	NUSEX * SASVABAI
INTER18	NUSEX * ADMINSCR
INTER19	TERMENLT * SASVABAI
INTER20	TERMENLT * ADMINSCR
INTER21	SASVABAI * ADMINSCR

meaningful. In this study, frequency analysis was performed to provide counts and percentage distributions of individuals in the sample, and to illustrate the range of the predictor and criterion variables. This information provided a base upon which to screen aberrant observations and to compare the results of this study. Frequency distributions are provided in Appendix B for the AD rating.

TABLE II

Selection Models

Model	Predictors/ Discriminating Variables		Regression Criterion Variable	Discriminant Classification Variable
A	DEPENDTS ADMINSCR BLACK NUSEX	TERMENLT HSDG OTHER	TAFMS1	SUCCTAF
B	TERMENLT INTER04 INTER14	INTER03 INTER08 INTER21	TAFMS1	SUCCTAF
C	INTER03 SASVABWK	INTER08 ENTRPAYG	TAFMS1	SUCCTAF
D	DEPENDTS OTHER SASVABAI	HSDG TERMENLT SASVABWK	SUCCESS2	SUCCESS2
E	INTER03	INTER09	SUCCESS2	SUCCESS2
F	INTER03 INTER21 SASVABEI SASVABSI CHYEC	INTER09 OTHER SASVABMK AFQTGRPS	SUCCESS2	SUCCESS2

Note: Variable sets A, B, D and E resulted from stepwise regression techniques. The variable sets C and F resulted from stepwise discriminant techniques.
Table I provides the labels for these variables.

2. Summary Statistics

Like frequency distributions, descriptive summary statistics are useful for analyzing and interpreting quantitative data. These summary statistics represent properties of location, dispersion and shape, and may be used to extract and summarize features of the data set. Representative summary statistics for variables in the AD data set are shown in Table III.

TABLE III

Selected Summary Statistics

VARIABLE	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE
AFQTPCNT	43.49	20.50	6.00	99.00
TAFMS1	49.22	9.44	2.00	72.00
ENTRYAGE	18.85	1.82	17.00	30.00
SASVAEGI	49.91	7.71	20.00	66.00
SASVAENO	50.43	7.80	23.00	69.00
SASVAEAD	50.34	9.28	20.00	80.00
SASVAEWK	48.20	7.51	30.00	64.00
SASVAEAR	48.92	6.98	29.00	65.00
SASVAESP	49.12	8.39	21.00	66.00
SASVAEMK	50.46	7.01	26.00	67.00
SASVAEEI	51.04	6.98	20.00	68.00
SASVAEMC	50.08	8.24	25.00	71.00
SASVAEGS	49.57	7.14	24.00	70.00
SASVAESI	50.90	8.48	20.00	65.00
SASVAEAI	51.16	9.51	26.00	67.00
CHYEC	11.79	0.73	3.50	16.00
ADCCMECS	199.99	19.19	99.00	264.00

3. Multivariate Correlation Analysis

Measuring the strength of the relationship between variables may be accomplished by correlation analysis. This technique enables one to gain an idea of the degree of association or covariation that a variable has with another variable. The summary measure that expresses the extent of this relationship is the coefficient of correlation, r, whose values range from -1 for perfect negative correlation to +1 for perfect positive correlation. Values close to zero indicate little systematic covariation between two variables. Correlation coefficients for quantitative variables used in this study are contained in Appendix B.

Assessing the strength of association between variables does not allow a researcher to predict the value of one variable from the value of another variable. The latter involves regression techniques, and is presented in the next section of this study.

E. REGRESSION ANALYSIS

Regression analysis is one method used to develop a statistical model that can predict the values of a dependent or response variable based on the values of independent or explanatory variables. Rather than merely measuring the association between variables with correlation analysis, a regression model attempts to predict or explain the value of the criterion variable by developing an equation that is based on weighted values of one or more predictor variables.

In developing the selection models in this study, the process employed was to first apply a variable "search" procedure called stepwise regression. The resultant models were then analyzed by simple regression analysis, and validated against a hold-out sample of the data set. The details of this process, the specific models derived, and results of the analysis are reported in the following sections. Appendix C contains a discussion of regression analysis assumptions and methodology.

1. Stepwise Regression

One of the desired characteristics of a regression model is parsimony, which means including the least number of explanatory variables that permit adequate interpretation of the dependent variable of interest. Such models are easier to interpret and are not as likely to be affected by multicollinearity[3] problems. In developing the models for this study, stepwise regression procedures were employed to find a "best" combination of predictor variables, thereby avoiding the computationally complex and costly process of examining all possible regressions.

---

[3]Multicollinearity refers to the condition in which some of the independent variables are highly correlated with each other. When multicollinearity is present, the values of the regression coefficients for the correlated variables may fluctuate dramatically.

In this study, two sets of candidate predictor variables were analyzed with the stepwise procedure. The first set included those entry-level attributes and measures that were considered likely to be good predictors of each criterion, based on a review of similar enlistment standards studies. As discussed in Chapter II, these variables included individual and demographic measures such as mental ability, amount of education, entry age, entry paygrade, marital status, AFQT percentile, and ASVAB scores. Table IV provides a list of the 18 candidate variables from the AD data set that were used in the stepwise procedure.

The second set of candidate predictor variables included the seven variables from the first set that met the specified significance level for inclusion in the stepwise model. In addition, this set included all two-level interactions[4] of these seven variables. Inclusion of interaction terms in this study represents a marked departure from previous enlistment standards research. The results of this analysis clearly indicate the presence of interaction effects among predictor variables. The seven predictor variables and 21 interaction terms used in the stepwise analysis are also contained in Table IV.

Using these two sets of candidate predictor variables, the stepwise procedure was run on each of the two criterion variables, TAFMS1 and SUCCESS2, which were defined in Chapter III. The resulting four models were developed from a uniform random split, the derivation sample, of 1440 observations in the AD data set. This derivation sample constituted approximately half of the 2820 total cases in the AD data set. So doing facilitated cross-validation of

---

[4]An interaction involves the product of two or more independent variables, and is included in a regression model when the relationship between one independent variable and the dependent variable changes for differing values of another independent variable [Ref. 11].

TABLE IV

Predictors in Stepwise Regressions

Variable	Label
	-- FIRST SET --
AFQTPCNT	- AFQT PERCENTILE (OR EQUIVALENT)
ENTRYAGE	- AGE OF INDIVIDUAL AT TIME OF ENTRY
ENTRPAYG	- ENTRY PAYGRADE (E0--O11)
TERMENIT	- TERM OF ENLISTMENT (NO. OF YEARS)
HSDG	- HIGH-SCHOOL GRADUATE(1) V. OTHER(0)
DEPENDTS	- SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
SASVABGI	- STANDARDIZED SCORE - GENERAL INFORMATION
SASVABNO	- STANDARDIZED SCORE -NUMERICAL OPERATIONS
SASVABAD	- STANDARDIZED SCORE - ATTENTION TO DETAIL
SASVABWK	- STANDARDIZED SCORE - WORD KNOWLEDGE
SASVABSP	- STANDARDIZED SCORE - SPACE PERCEPTION
SASVABMC	- STANDARDIZED SCORE - MECH COMPREHENSION
SASVABSI	- STANDARDIZED SCORE - SHOP INFORMATION
SASVABAI	- STANDARDIZED SCORE - AUTO INFORMATION
BLACK	- (1) BLACK, ELSE (0)
OTHER	- (1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX	- (1) MALE, (0) FEMALE
ADMINSCR	- AD ASVAB COMPOSITE SCREEN
	-- SECOND SET --
TERMENIT	- TERM OF ENLISTMENT (NO. OF YEARS)
HSDG	- HIGH-SCHOOL GRADUATE(1) V. OTHER(0)
DEPENDTS	- SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
SASVABAI	- STANDARDIZED SCORE - AUTO INFORMATION
BLACK	- (1) BLACK, ELSE (0)
NUSEX	- (1) MALE, (0) FEMALE
ADMINSCR	- AD ASVAB COMPOSITE SCREEN
INTER01	- DEPENDTS * HSDG
INTER02	- DEPENDTS * BLACK
INTER03	- DEPENDTS * NUSEX
INTER04	- DEPENDTS * TERMENLT
INTER05	- DEPENDTS * SASVABAI
INTER06	- DEPENDTS * ADMINSCR
INTER07	- HSDG * BLACK
INTER08	- HSDG * NUSEX
INTER09	- HSDG * TERMENIT
INTER10	- HSDG * SASVABAI
INTER11	- HSDG * ADMINSCR
INTER12	- BLACK * NUSEX
INTER13	- BLACK * TERMENIT
INTER14	- BLACK * SASVABAI
INTER15	- BLACK * ADMINSCR
INTER16	- NUSEX * TERMENIT
INTER17	- NUSEX * SASVABAI
INTER18	- NUSEX * ADMINSCR
INTER19	- TERMENLT * SASVABAI
INTER20	- TERMENIT * ADMINSCR
INTER21	- SASVABAI * ADMINSCR

the models against a hold-out sample, the validation sample, whose characteristics would not influence the original

development of the models. The predictor variables that remained in the model at the termination of the stepwise procedure were significant at $p < .10$, and most variables were significant at $p < .05$. The four models themselves were significant at $p < .0001$.

2. Multiple Regression

The four models developed by the stepwise process were next analyzed using the SAS Regression procedure to describe the particular straight line model that best fit the data. Table V contains the printed output from the SAS Regression procedure that was run on each of the four models. For comparative purposes, two models developed by discriminant analysis techniques, discussed in the next section of this chapter, are also shown in table V. The SAS User's Guide provides a detailed description of the statistics that are included in the tables, as well as their method of computation [Ref. 12]. It can be seen that Model E, with the highest R-SQUARE and all variables statistically significant, is the preferred regression model.

The proportion of variation in the criterion variable explained by the set of predictor variables selected is called the coefficient of multiple determination, and is denoted R-SQUARE. The values of R-SQUARE for the models developed in this study are relatively low. This may be partially attributable to the large number of observations in the AD data set. However, it is also likely that the variation of the criterion variable, length of service or success as defined in this study, is also due to factors not captured by the entry-level attributes and measures used as predictor variables. These factors, which affect an individual's performance and decision to remain in the service, present themselves subsequent to enlistment. They may include satisfaction with initial assignment, geographical

TABLE V

Regression Analysis Results

Model	Predictors	Parameter Estimates	Prob > |T|	R Square	F Value
A	INTERCEPT	29.049	0.0001	0.0537	11.613
	DEPENDTS	2.841	0.0636		
	TERMENLT	3.639	0.0001		
	ADMINSCR	-1.207	0.0260		
	HSDG	1.807	0.0036		
	CTEER	2.254	0.0294		
	MCSEX	4.171	0.0079		
	BLACK	1.729	0.0131		
B	INTERCEPT	32.140	0.0001	0.0547	13.828
	TERMENLT	3.890	0.0001		
	INTER03	15.724	0.0026		
	INTER04	-2.937	0.0173		
	INTER08	2.113	0.0004		
	INTER14	0.032	0.0398		
	INTER21	-0.024	0.0134		
C	INTERCEPT	51.746	0.0001	0.0220	8.089
	INTER03	3.888	0.0163		
	INTER08	2.137	0.0004		
	SASVABWK	-0.101	0.0022		
	ENTLPAYG	0.416	0.3685		
D	INTERCEPT	0.535	0.0002	0.0255	6.238
	DEPENDTS	0.172	0.0131		
	TERMENLT	0.053	0.0549		
	HSDG	0.115	0.0001		
	CTEER	0.080	0.0871		
	SASVABAI	0.001	0.5630		
	SASVABWK	-0.003	0.1028		
E	INTERCEPT	0.663	0.0001	0.0198	14.501
	INTER03	0.196	0.0064		
	INTER09	0.030	0.0001		
F	INTERCEPT	0.565	0.0309	0.0370	6.107
	INTER03	0.202	0.0053		
	INTER09	0.038	0.0001		
	INTER21	-0.001	0.0576		
	CTEER	0.101	0.0297		
	SASVABEI	0.006	0.0022		
	SASVABSI	0.002	0.1456		
	CHYEC	-0.033	0.1138		
	AFQTGRPS	-0.027	0.0092		

location of duty assignment, command climate, unit employment, change in marital status, societal values and pressures, and educational and economic opportunities outside the military. These factors or measures are post hoc considerations that are not available when screening candidates for enlistment and initial rating assignment. They

are issues that are appropriate for more sophisticated methodologies, for example, covariance structure analysis which can treat complicated enlistment standards models as a series of simultaneous equations that capture performance as a "multiple-stage" process occurring throughout an individual's military career. [Ref. 10]

3. Validation

The results of the regression procedure were next validated against the hold-out sample. Each of the regression models was derived from a uniform random sample, the derivation sample, of the observations in the AD data set. The SAS Regression procedure was employed to calculate the parameter estimates for the associated predictor variables using data from observations in this derivation sample. The SAS Score procedure then used these estimates to predict the value of the criterion variable for each observation in the validation sample. Finally, these predicted values were correlated with the actual values of the criterion in the validation sample. These correlations represent the validation coefficients for each model, and are shown in Table VI.

TABLE VI

Regression Model Validities

Model	First Validity Coefficient	Second Validity Coefficient	Average Validity
A	0.21342	0.20317	0.21
B	0.21536	0.21683	0.22
C	0.14459	0.13612	0.14
D	0.17387	0.13766	0.16
E	0.17790	0.12751	0.14
F	0.14430	0.13531	0.14

Note: The First Validity Coefficient is the result of the cross-validation, and the Second Validity Coefficient results from the double cross-validation. The reported average is the simple arithmetic mean.

As a further check of the validity of the six regression models, the process was repeated by deriving parameter estimates from the validation sample, and using these estimates to correlate the actual and predicted values of the criterion for observations in the derivation sample. This double cross-validation technique is described in detail by Campbell [Ref. 13]. Table VI also contains this second set of validity coefficients for the six models.

Occasionally, concern is expressed that random samples may not be from a homogeneous population, and, therefore, the sample correlations may differ from the population correlations. One method of addressing the problem of heterogeneous samples is to average the correlation coefficients to obtain a single estimate of the population correlation. If the sample correlations are of about the same value and if they are not too large, as is the case with this study, this simple arithmetic mean will suffice. Were this not the case, however, another technique is to use transformations to Fisher's z coefficients. [Ref. 14] The simple arithmetic average correlations are also presented in Table VI. Appendix C contains the program used to calculate validity coefficients.

C. DISCRIMINANT ANALYSIS

The third statistical method employed in this research was discriminant analysis. The use of discriminant analysis allows observations to be classified into two or more groups or categories on the basis of one or more numeric variables. As was done with regression analysis, the discriminant models were derived and analyzed from the derivation sample of the data set, and tested against the hold-out sample of observations. Variables in the model were again selected using stepwise techniques. The resulting two models, and

the four models developed by regression analysis, were then analyzed using the SAS Discriminant procedure. The program used in this analysis is contained in Appendix D, along with a discussion of discriminant analysis assumptions and methodology.

1. Stepwise Discriminant Analysis

The SAS Stepwise Discriminant procedure was employed to select the most useful discriminating variables. It is a logical and efficient method of choosing an optimal combination of variables. Their selection to enter or leave the model is based on either the significance level of an F test or a squared partial correlation criterion. The selected variables are those which contribute most to the discriminatory power of the model. [Ref. 12]

The variables chosen by the stepwise discriminant process were selected from the 46 candidate variables shown previously in Table I. The entry-level attributes and measures that were considered likely to be good predictors, as discussed in Chapter II, represent 25 of these candidate variables. The other 21 variables are the two-level interaction terms considered during regression analysis of the AD data set. The procedure was run on each of the two criterion variables, SUCC1AF and SUCCESS2, discussed in Chapter III. The criterion variables define the groups into which each observation will be classified by discriminant analysis, and are called classification variables.

2. Discriminant Analysis

As previously mentioned, discriminant analysis involves the study of differences between two or more groups, defined by a single nominal level variable, with a set of common discriminating variables.

The SAS Discriminant Analysis procedure provided the means for conducting discriminant analysis of the AD data set. The procedure was run on each of the six models developed by stepwise regression and stepwise discriminant processes. Each observation is placed in the class from which it has the smallest generalized squared distance. Also taken into account were the prior probabilities of group membership. These probabilities are obtained from a frequency distribution of actual successes and failures of the sample data set. This was considered appropriate since this study is attempting to improve upon the selection process in use at the time the individuals enlisted.

Table VII contains the results of discriminant analysis. Each procedure incorporated the prior probability of group membership, indicated on the classification matrix as PRIORS. The classification matrix is divided into four elements which depict the number of actual (row) versus predicted (column) classifications into successful (1) or failure (0) groups. The four elements (actual, predicted) in the matrix are:

(0,0) The number of failure cases predicted to be failures

(1,0) The number of successful cases predicted to be failures

(0,1) The number of failure cases predicted to be successful

(1,1) The number of successful cases predicted to be successful

Each section first contains the classification matrix developed by applying the classification function to the derivation sample. The second classification matrix depicts the results of applying this same classification function to observations in the hold-out sample, thereby validating the model.

The table also shows two rates relevant to each classification matrix. The first rate is the percentage of correct classifications, called the "hit rate", which provides a measure of the accuracy of the discriminant model. The second rate is the percentage of enlistees who were classified as (1,1) compared to all cases who were predicted as successful. It is called the "success rate", and it provides a measure of how well this selection model would have performed. It may be compared to the original selection strategy success rate, the priors. Success rate is an important consideration with utility analysis, and will be addressed further in Chapter V. As with regression analysis, Model B is again the preferred model since it is the only one that improves upon the selection strategy in existence during the timeframe of the AD data set.

To illustrate how the results may be interpreted, an example of the classification matrices for Model A will be explained. The model correctly classified 49 observations as failures and 1079 observations as successful. The sum of these correct classifications represents 78 percent of the total of 1440 observations in the derivation sample. To test the model's accuracy, the classification function is applied to the validation sample. The second classification matrix indicates 47 failure and 1039 successful observations were correctly classified. The sum represents a hit-rate of 79 percent of the total of 1380 observations in the hold-out sample. The consistency of these hit-rates indicates the model is valid. The model betters the 85 percent success rate experienced by the Navy with the selection process used at the time the observations enlisted.

However, it is difficult to significantly improve upon such a high success rate. Additional entry-level attributes and measures might be found to better capture success as defined in this study. An alternate approach

would be to redefine the success criterion. In either case, however, the methodology presented in this chapter may be similarly followed to develop and test enlistment standards models. The next chapter will discuss a method by which the utility of such an effort may be measured.

TABLE VII

Discriminant Analysis Results

Model	Priors 0	Priors 1	Classification Matrix	Actual SUCCIAF	Predicted SUCCIAF 0	Predicted SUCCIAF 1	Total	Hit Rate	Success Rate
A	0.15	0.85	Actual SUCCIAF	0	49	161	210	0.78	0.87
				1	151	1079	1230		
			Total		200	1240	1440		
			Actual SUCCIAF	0	47	136	183	0.79	0.88
				1	158	1039	1197		
			Total		205	1175	1380		
B	0.15	0.85	Actual SUCCIAF	0	1	209	210	0.85	0.85
				1	2	1228	1230		
			Total		3	1437	1440		
			Actual SUCCIAF	0	0	183	183	0.87	0.87
				1	1	1196	1197		
			Total		1	1379	1380		

Model	Errors 0	Errors 1	Classification Matrix					Hit Rate	Success Rate
C	0.15	0.85			Predicted SUCCTAF			0.83	0.86
					0	1	Total		
			Actual SUCCTAF	0	15	195	210		
				1	46	1184	1230		
			Total		61	1379	1440		
					Predicted SUCCTAF			0.83	0.87
					0	1	Total		
			Actual SUCCTAF	0	8	175	183		
				1	62	1135	1197		
			Total		70	1310	1380		
D	0.23	0.77			Predicted SUCCESS2			0.36	0.86
					0	1	Total		
			Actual SUCCESS2	0	302	35	337		
				1	889	214	1103		
			Total		1191	249	1440		
					Predicted SUCCESS2			0.35	0.84
					0	1	Total		
			Actual SUCCESS2	0	277	41	318		
				1	850	212	1062		
			Total		1127	253	1380		

Model	Priors 0	Priors 1	Classification Matrix		Predicted SUCCESS2 0	Predicted SUCCESS2 1	Total	Hit Rate	Success Rate
E	0.23	0.77			0	1	Total	0.70	0.79
			Actual SUCCESS2	0	95	242	337		
				1	187	916	1103		
			Total		282	1158	1440		
					0	1	Total	0.72	0.79
			Actual SUCCESS2	0	112	206	318		
				1	174	888	1062		
			Total		286	1094	1380		
F	0.23	0.77			0	1	Total	0.55	0.85
			Actual SUCCESS2	0	238	99	337		
				1	554	549	1103		
			Total		792	648	1440		
					0	1	Total	0.50	0.82
			Actual SUCCESS2	0	218	100	318		
				1	591	471	1062		
			Total		809	571	1380		

V. UTILITY ANALYSIS

This chapter contains an explanation of the applicability of utility analysis to the development of selection procedures, and discusses the theory of utility analysis. The methodology used in this study to apply utility analysis is described, along with sections on the calculation of cell probablities for regression and discriminant models, and a section on estimating cell utilities. More detail on the calculations and programs used for utility analysis may be found in Appendix E.

A. PURPOSE OF UTILITY ANALYSIS

The development of a model for use in predicting an applicant's future performance in a particular job is a very necessary part of most selection procedures. However, the model itself does not constitute enough information to enable a decision to be made on whether or not it is worth implementing. The validity of the model is one indicator of its potential usefulness but, as will be seen, other factors significantly affect the usefulness of a model. All organizations would find it valuable to be able to judge the worth of their strategy in quantitative terms, particularly when comparing their existing strategy to a newly developed, competing strategy. A framework is needed which will allow the evaluation of a selection model in terms of the institutional gains (or losses) that are expected to result when that model is used to guide decisions on selection. Classical utility analysis provides such a framework, and it allows the calculation of usefulness to be made in terms of actual dollars, which facilitates the comparison of one selection model with another.

E. THEORY OF UTILITY ANALYSIS

In the context of utility analysis, there are four outcomes of interest associated with selection decisions. These outcomes are:

- Valid Positives (VP), which refers to the number of applicants that are hired and who turn out to be successful on the job.
- False Positives (FP), which refers to the number of applicants that are hired and who turn out to be unsuccessful on the job.
- False Negatives (FN) are the people who were not hired, but who would have been successful if they had been hired.
- Valid Negatives (VN) are the people that were not hired, and who would have been unsuccessful if they had been hired.

It is obvious from the terminology and the explanations that VP and VN constitute correct selection decisions, and FP and FN represent selection error.

These outcomes are perhaps easier to understand with the aid of a diagram. Figure 5.1 shows the relationship between hypothetical predicted (from a model) and actual scores on a job performance criterion for a large number of job applicants.

The ellipse contains the data on predicted and actual criterion scores. In this diagrammatic example, the correlation between the predicted and actual scores (the model's validity) is apparent--higher predicted scores are associated with higher actual scores and vice versa. The point y on the vertical axis is the dividing line between what is considered to be successful performance (say completion of 48 months of service for first term enlistees), and unsuccessful performance (less than 48 months service before

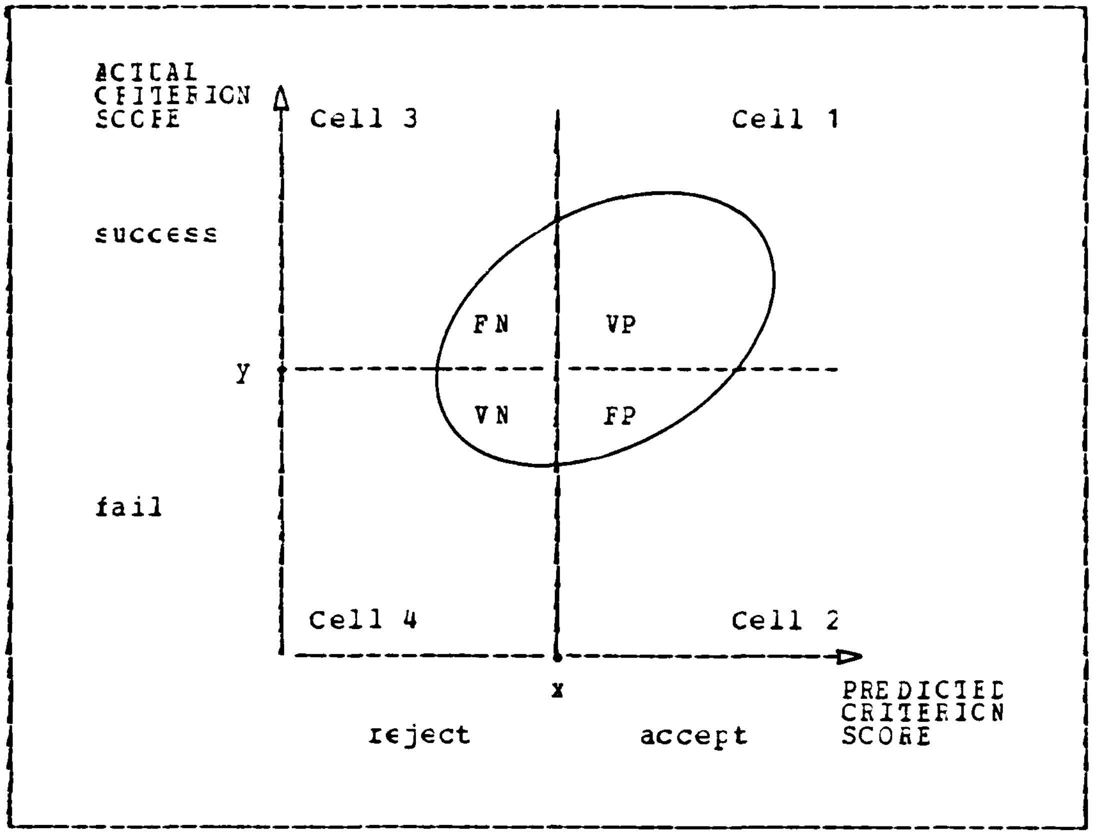


Figure 5.1 Hypothetical Predicted and Actual Scores

discharge). In utility analysis the term _base rate_ is defined as that proportion of current employees who are considered to be successful. If seven out of every ten employees are successful, then the base rate is .70. The point _x_ on the horizontal axis is referred to as the _cut score_. If an applicant's predicted score (from the model) is greater than _x_, then that person will be accepted (hired), and if their predicted score is less than _x_, then they will be rejected (not hired). The location of _x_ on the horizontal axis will often depend on the _selection ratio_, which is the proportion of applicants that need to be accepted in order to fill a certain number of jobs. If,

over the course of one year, 80 job vacancies are expected to occur and if 100 applicants over the year are expected to apply for those jobs, then the selection ratio needs to be .80 if all vacancies are to be filled. In the happy circumstance (from the recruiter's point of view) where there are far more applicants than jobs, then the cut score _x_ will be chosen so as to maximize the _utility_ of the selection procedure. Utility is defined here to mean the expected gain in dollars that results from a particular selection strategy.

The lines generated from the base rate and the cut score divide the sample into four cells as shown. Each cell contains the people who are classified into each of the four outcomes of interest. In cells 1 and 2 are people whose predicted score is higher than the cut score. Therefore these people would be classified as accept. These accepted people (the positives) are further divided into those who would be successful (valid positives) and those who would be unsuccessful (false positives). Cells 3 and 4 contain the people who scored lower than the cut score on the predictor, and these would be classified as reject. Again, some of these rejected cases would have been successful (false negatives), and some would have failed (valid negatives). In utility analysis it is convenient to convert the cell counts (represented by VP, FP, FN and VN) to proportions of the overall sample, so each count is divided by the number of people in the sample and the cell probabilities (PVP, PFP, PFN and PVN respectively) result.

One further result of interest is the _success rate_. The success rate is defined as the proportion of hired applicants who are, or will be, successful. It is simply found by dividing PVP by the sum of PVP and PFP.

Given the concepts and terminology outlined above, it is now possible to discuss in general terms the factors that will affect the cell probabilities which, in turn, affect the expected utility.

1. Model Validity

The model's validity, as measured by the correlation between predicted and actual scores, is one factor that determines the degree of selection error resulting from the selection strategy. If the validity is high, then the proportions of correctly classified people (PVP and PVN) will be higher, and the selection error (PFP and PFN) will be lower. Vineberg and Joyner in their review of almost 150 military studies related to job performance prediction, found that validities range from .15 to .40, from a total of 350 validity coefficients [Ref. 15]. Generally, validities within this range would be considered as low or medium.

2. Base Rate

If the existing base rate is high (say .70 or greater), then it means that whatever selection strategy is currently in use has a high rate of success in identifying potentially successful applicants. Under these circumstances, it is unlikely that using a new model in the staffing process would yield much of an improvement in correctly selecting applicants. A high base rate means that the cell probabilities for PVP and PFN are going to be higher than for PFP and PVN.

3. Selection Ratio

Assuming the model is valid, the lower the selection ratio, the more useful the model will be in identifying successful applicants. Decreasing selection ratios mean that the organization can be increasingly selective in whom it hires. Naturally, it will tend to accept only those who score highest on the predictor, those who are also predicted most likely to be successful. A low selection ratio (high cut score) will mean that PVP and PFP will be small. It

also follows that a low selection ratio will yield a higher success rate--although few people will be hired, most of them will represent correct selection decisions (PVP).

C. ESTIMATING THE UTILITY OF A MODEL

The expected utility (EU) of a model is found by summing the products of each cell probability and its associated cell utility (U1, U2, U3 and U4), and subtracting the cost of giving the test to an applicant (UT).

$$EU = U1(PVP) + U2(PFP) + U3(PFN) + U4(PVN) - UT \qquad (5.1)$$

Appendix E contains detailed descriptions on how cell probabilities and cell utilities are determined. For a discriminant model the cell probabilities may be readily derived from the output of the SAS Discriminant procedure, because the model classifies cases into predicted successes and predicted failures. In the regression model the cut score is not known in advance, so cell probabilities that result from a number of possible cut scores are calculated, and a cut score is eventually chosen based on which set of cell probabilities maximizes the utility of the model.

The formula for calculating the expected utility of a model requires that a utility be assigned to each selection outcome. These cell utilities are designated U1 through U4 and are associated with the outcomes VP, FP, FN and VN respectively. The Billet Cost Model provides an estimate of the cost to the Navy of staffing a billet. In this study it is assumed that this cost is equal to the marginal product of a successful sailor, and so the utility of a valid positive (U1) is assigned a value of $24,163 [Ref. 16]. No proven technique exists for estimating the cell utilities for the three other selection outcomes. Individual

circumstances and prevailing market conditions make it difficult to estimate these outcomes with real confidence, so these cell utilities were estimated relative to U1, and a minor form of sensitivity analysis was conducted. The cell utility of a false positive (U2) was assigned values of -.5, -1 and -2. Valid negatives (U4) were assigned an equal and opposite utility to U2, and false negatives (U3) were assigned values of 0, -.25 and -.5. Table VIII shows seven different sets of cell utilities that were considered.

TABLE VIII

Relative Cell Utilities

U1	U2	U3	U4
1.0	-0.5	-0.25	0.5
1.0	-0.5	0.0	0.5
1.0	-1.0	0.0	1.0
1.0	-2.0	0.0	2.0
1.0	-0.5	-0.5	0.5
1.0	-1.0	-0.5	1.0
1.0	-2.0	-0.5	2.0

The cost of administering a test (UT) is of significance if the costs of testing are different for competing selection strategies. The models developed in this study use data gathered from the existing tests, and therefore the costs of testing will remain much the same. Thus in this context, UT may be ignored since it applies equally to the old and new tests.

D. RESULTS OF UTILITY ANALYSIS

Tables IX and X contain the results of the utility analysis on the regression and discriminant models respectively. The "Percent Change in EU" column is the result of the comparison of the model's utility with the utility of the

Navy's original selection strategy (base line utility). A positive percentage change in EU indicates that the maximum utility obtainable from the model is higher than the utility of the original selection strategy. An increase in utility of say $50 means that the Navy saves $50 for each selection decision (correct or incorrect) that is made by using the model rather than the original strategy. For the models with the SUCCTAF or TAFMS1 criterion, the base rate is .861, i.e., 86.1 percent of the people selected by the Navy were successful. These people can be thought of as the valid positives of the original strategy and the remaining 13.9 percent are false positives. (For the SUCCESS2 criterion these figures are 76.8 percent and 23.3 percent.) Unfortunately it is not possible to calculate the values of false and valid negatives so these are considered to be zero. For the TAFMS1 or SUCCTAF criterion then, the cell probabilities for the original selection strategy are PVP = .861, PFP = .139, PFN = 0 and PVN =0. The base line utility for each of the three different combinations of U1 and U2 can then be calculated. The model utilities are then compared to these base line utilities and the differences, expressed as a percentage of the base line utilities, are reported. Similarly the base success rate of the original strategy is also .861 (for the TAFMS1 or SUCCTAF criterion). The column "Change in Succrate" reports the actual difference between the models' success rates and the base success rates. The column "SRATIO" shows the selection ratio that results when the cut score is chosen so as to maximize the utility, for each set of cell utilities.

1. Regression Models

For most sets of cell utilities, the regression models developed show little improvement over the original selection strategy. In most cases the selection ratio is

very close to 1 and the percentage increase in expected utility is very small. This is not a surprising result because the model validities are relatively low (around .20) and, more significantly, the base rates are very high at .861 and .768. It is interesting to note, however, that when the costs of a false positive and the benefits of a valid negative are high, then the selection ratio is driven down, and the utility and success rate go up.

2. Discriminant Models

In general the discriminant models did not perform as well as the regression models or the Navy's original selection strategy. For some models the percent change in EU was a significant positive number, but these were usually associated with extreme assumptions of cell utilities. In addition to the factors mentioned in the previous subsection, this poor performance is because the discriminant models lack the flexibility to vary the cell probabilities depending on the values of the cell utilities. There is no option to vary predictions depending on the consequences of correct and incorrect selection decisions, and thus only one set of cell probabilities is available for each discriminant model.

TABLE IX

Utility Results - Regression Models

MODEL	U1	U2	U3	U4	Δ% EU	ΔSUCCRATE	SRATIO
A	1.0	-0.5	-0.25	0.5	0.12	0.001	0.998
	1.0	-0.5	0	0.5	0.14	0.001	0.998
	1.0	-1.0	0	1.0	0.34	0.001	0.998
	1.0	-2.0	0	2.0	5.85	0.022	0.805
	1.0	-0.5	-0.5	0.5	0.11	0.001	0.998
	1.0	-1.0	-0.5	1.0	0.32	0.001	0.998
	1.0	-2.0	-0.5	2.0	1.25	0.003	0.998
B	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
	1.0	-1.0	0	1.0	0.0	0.0	1.0
	1.0	-2.0	0	2.0	6.28	0.023	0.8 10
	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	0.28	0.002	0.985
C	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
	1.0	-1.0	0	1.0	0.05	0.001	0.998
	1.0	-2.0	0	2.0	5.79	0.016	0.871
	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	0.4	0.004	0.972
D	1.0	-0.5	-0.25	0.5	0.15	0.002	0.995
	1.0	-0.5	0	0.5	0.22	0.002	0.995
	1.0	-1.0	0	1.0	5.1	0.027	0.861
	1.0	-2.0	0	2.0	72.98	0.074	0.283
	1.0	-0.5	-0.5	0.5	0.08	0.002	0.995
	1.0	-1.0	-0.5	1.0	0.76	0.002	0.995
	1.0	-2.0	-0.5	2.0	35.44	0.014	0.806
E	1.0	-0.5	-0.25	0.5	0.0	0.0	1.0
	1.0	-0.5	0	0.5	0.0	0.0	1.0
	1.0	-1.0	0	1.0	3.51	0.033	0.799
	1.0	-2.0	0	2.0	61.76	0.124	0.056
	1.0	-0.5	-0.5	0.5	0.0	0.0	1.0
	1.0	-1.0	-0.5	1.0	0.0	0.0	1.0
	1.0	-2.0	-0.5	2.0	33.51	0.033	0.799
F	1.0	-0.5	-0.25	0.5	0.14	0.001	0.997
	1.0	-0.5	0	0.5	0.16	0.001	0.997
	1.0	-1.0	0	1.0	4.77	0.013	0.816
	1.0	-2.0	0	2.0	79.18	0.063	0.511
	1.0	-0.5	-0.5	0.5	0.11	0.001	0.997
	1.0	-1.0	-0.5	1.0	0.46	0.001	0.997
	1.0	-2.0	-0.5	2.0	36.61	0.014	0.807

Note: The base utilities for Models A, B and C are -$19112 (when U2 is -0.5), $17428 (when U2 is -1.0) and $14061 (when U2 is -2), and the base success rate is 0.861.

The base utilities for Models D, E and F are -$15744 (when U2 is -0.5), $12938 (when U2 is -1.0) and $7326 (when U2 is -2.0), and the base success rate is 0.768.

TABLE X

Utility Results - Discriminant Models

MODEL	U1	U2	U3	U4	Δ% EU	Δ SUCCRATE	STATIC
A	1.0	-0.5	-0.25	0.5	-13.0	0.016	0.856
	1.0	-0.5	0	0.5	-9.5		
	1.0	-1.0	0	1.0	-5.8		
	1.0	-2.0	0	2.0	4.6		
	1.0	-0.5	-0.5	0.5	-16.5		
	1.0	-1.0	-0.5	1.0	-13.3		
	1.0	-2.0	-0.5	2.0	-4.8		
B	1.0	-0.5	-0.25	0.5	-0.1	0.0	0.999
	1.0	-0.5	0	0.5	-0.1		
	1.0	-1.0	0	1.0	-0.1		
	1.0	-2.0	0	2.0	0.1		
	1.0	-0.5	-0.5	0.5	-0.2		
	1.0	-1.0	-0.5	1.0	-0.1		
	1.0	-2.0	-0.5	2.0	0.0		
C	1.0	-0.5	-0.25	0.5	-5.0	0.074	0.954
	1.0	-0.5	0	0.5	-3.8		
	1.0	-1.0	0	1.0	-3.1		
	1.0	-2.0	0	2.0	-1.0		
	1.0	-0.5	-0.5	0.5	-6.2		
	1.0	-1.0	-0.5	1.0	-5.7		
	1.0	-2.0	-0.5	2.0	-4.3		
D	1.0	-0.5	-0.25	0.5	-86.8	0.081	0.178
	1.0	-0.5	0	0.5	-63.1		
	1.0	-1.0	0	1.0	-38.5		
	1.0	-2.0	0	2.0	67.5		
	1.0	-0.5	-0.5	0.5	-110.0		
	1.0	-1.0	-0.5	1.0	-96.8		
	1.0	-2.0	-0.5	2.0	-25.3		
E	1.0	-0.5	-0.25	0.5	-13.3	0.033	0.799
	1.0	-0.5	0	0.5	-8.4		
	1.0	-1.0	0	1.0	3.5		
	1.0	-2.0	0	2.0	54.6		
	1.0	-0.5	-0.5	0.5	-18.2		
	1.0	-1.0	-0.5	1.0	-8.4		
	1.0	-2.0	-0.5	2.0	33.5		
F	1.0	-0.5	-0.25	0.5	-53.1	0.069	0.432
	1.0	-0.5	0	0.5	-37.5		
	1.0	-1.0	0	1.0	-15.4		
	1.0	-2.0	0	2.0	79.4		
	1.0	-0.5	-0.5	0.5	-68.7		
	1.0	-1.0	-0.5	1.0	-53.3		
	1.0	-2.0	-0.5	2.0	12.5		

Note: The base utilities for Models A, B and C are -\$19112 (when U2 is -0.5), \$17428 (when U2 is -1.0) and \$14061 (when U2 is -2), and the base success rate is 0.861.

The base utilities for Models D, E and F are -\$15744 (when U2 is -0.5), \$12938 (when U2 is -1.0) and \$7326 (when U2 is -2.0), and the base success rate is 0.768.

VI. CONCLUSIONS AND RECOMMENDATIONS

This study set out to provide a method for developing enlistment standards models which improves upon similar processes presently in use. Toward that end, significant advances have been made, particularly when compared to prior studies conducted at the Naval Postgraduate School. The techniques used provide a much more comprehensive approach to model development. They employ regression analysis to fully develop the stepwise regression results. In addition, stepwise discriminant procedures were used to find an optimal model prior to full discriminant analysis. Alternative criteria for measuring successful operational performance, including a continuous length of service criterion, were incorporated in the models. Finally, each model was analyzed using both regression and discriminant analysis techniques.

Perhaps <u>most</u> significant is the presentation of a means by which the benefits from such efforts may be gauged. The development of innovative utility analysis programs affords future researchers an excellent opportunity to measure in monetary terms the benefits to be derived from implementing a new selection strategy. It is important to reiterate that the statistical and utility analysis techniques presented in this study may be easily applied or modified to accommodate selection standards model development for any of the more than 90 Navy ratings contained in the master data base.

A secondary purpose of this study was to discover whether the models developed improve upon existing selection and assignment strategy for the AD rating. By and large, the models presented do not appreciably enhance the processes used since 1976. The models do, however, allow

one to focus on some specific considerations in the current screening processes. For example, Models A, B, and C allow policy makers to consider length of service in months, and to vary the criterion for measuring success. This capability is particularly appropriate for use in a dynamic recruiting market.

A. RESULTS

This study yielded several other results worth noting. The term of enlistment variable may be used to predict success now that it has been corrected to reflect active duty obligation. This is particularly important when assessing Naval Reservists, whose six year contract generally requires only three years of active service. The change from a negative to a significantly positive correlation of TERMENLT on the criteria is one of the more important discoveries of this research effort.

This study also determined that the usefulness of the SCREEN composite score in predicting job performance measures was virtually nonexistent. It appears to be more appropriate to use the SCREEN score components in the models, at least when attempting to predict operational job performance. Nontraditional ASVAB subtest scores, such as Auto Information, may also be appropriate for use in the screening process. Another significant finding of this study is the definite presence of interaction effects. Considering personal measures on an individual in conjunction with other measures represents a marked change in current selection practices.

To summarize the results of the statistical analysis, the variables measuring term of enlistment, education, dependents status, sex and race emerged as repeatedly significant predictors of successful operational

performance. The composite measure of eligibility for the AD rating, and the ASVAB Auto Information subtest score, were also significant predictor variables. In addition, Model E was shown to be the best regression and discriminant model.

The results of the application of utility analysis show that the regression models developed in this study perform as well as or better than the original Navy strategy which was used as the comparison (base line utility). It is important to note however, that the methodology used in this part of the study ensured that regression models will provide a maximum utility <u>at least</u> equal to the base line utility. This is because the technique allows the cut score to be set so low that all cases are accepted. Models A and F are considered to be the best of the models because they provide for significant increases in utility without having to resort to impractically low selection ratios. The discriminant Models A and E are better than the others because improvement over base utility is possible, depending on the cell utilities.

As was mentioned in Chapter V, the high existing base rates are an indication that newly developed models are unlikely to produce superior results. Utility analysis is hindered by the difficulty of confidently estimating the individual cell utilities, and this is an area that is in need of further research. It is also difficult to compare new selection strategies to existing ones because it is impossible to classify the cases rejected by the existing strategy as valid or false negatives. Data of this sort can only be obtained by testing all applicants and then accepting all of them, regardless of their relationship to the cut score, or to the desired selection ratio.

E. RECOMMENDATIONS

Despite the advances made by this study, there remains many opportunities to refine the models presented for the AD rating, and to develop models for other Navy ratings. Procedurally, these opportunities include testing for curvilinearity of the models, expanding the interaction terms to three or more levels, and seeking different combinations of ASVAB subtest scores as potential predictors. There may also be other measures not evaluated by this study that are significant operational performance predictors, such as enlistment waivers, DEP status, or involvement with civil authorities.

Consideration should also be given to altering the criterion variables. One particularly promising adjustment may be to change the criterion to reflect achieving E-5. This may be appropriate since the models developed appear to do a better job of predicting longer LOS, as indicated by preliminary residual analysis. Developing separate models that yield predictions of shorter LOS may also be in order.

The multiple-stage analytic approach referred to in Chapter IV also appears to be a promising technique. Such analysis might consider change in dependent status, performance evaluations, or advancement exam results as variables in a model.

To improve the usefulness of utility analysis it is important that a technique be developed to estimate cell utilities with reasonable accuracy. Such a technique needs to be able to control for changes in the recruiting market, and be sensitive to the changing Navy requirements for recruits. It is also important that data be gathered on applicants who are not accepted into a particular rating, to allow researchers to determine if they were reclassified to another rating, or rejected entirely.

In conclusion, it is clear that continued efforts to develop selection standards models for all ratings are essential. For it is through these efforts that the cost of training and maintaining Navy personnel will be reduced. The resultant experienced career force will ensure the Navy is ready to meet any global commitment.

APPENDIX A

DATA BASE DEVELOPMENT PROGRAMS

This appendix provides the SAS programs used in this study to access the master data base, develop the AD data set, and create new predictor and criterion variables, as discussed in Chapter III. Each program contains the job control language information appropriate to the Naval Postgraduate School's IBM 3033 computer system. Statistical Analysis System (SAS) statements are employed in the programs to accomplish the desired functions. These SAS statements are normally preceded by comments to explain their purpose, the comments being identified by an asterisk.

Table XI contains the program called "ADSETUP". This program was used to access the master file and extract information on Aviation Machinist's Mates (AD). (The master file tape, originally called "ENLIST", has recently been revised and relabeled "NPS709".) The data file created by this program is called "ADDATA", and it contains the initial 243 variables from the master file. Also provided in the program are the variable names and labels. The program may be used to extract data from the master file for any of approximately 90 Navy ratings simply by entering the appropriate abbreviation and four digit code for the selected rating.

Table XII provides the program called "ADSCREEN" that was used to screen the data extracted from the master file. These screens were performed on observations in the "ADDATA" file, and the results were placed in a file called "ADSCRSET". Because of the large number of cases and variables in the data, sufficient computing work space was not available. Therefore, the SAS KEEP statement was used to

retain 116 of the initial variables for analysis. It was felt these 116 variables captured all the desired measures on the observations that would be required for analysis. The last screen was incorporated following frequency distribution analysis to remove cases that had aberrant or impossible data associated with them.

Table XIII contains the program called "ADNEWVAR". This program was employed to create new predictor and criterion variables, as discussed in Chapter III. The program used information on observations in the "ADSUBSET" data file to create the new variables, and placed the results of these operations in a file called "ADALL4". This file thus constitutes the AD data set referred to throughout this study. It contains all of the selected and created variables that provide information on the 2820 ADs who remained in the data set after all screens were accomplished. It is this file that was used to conduct the statistical analysis for this study.

The "ADNEWVAR" program lists all created variable names and labels. It also contains the SAS statements that converted several qualitative variables to numeric variables or dichotomous (0,1) variables. Finally, the program shows the SAS statement used to split the AD data set into the two uniformly distributed random samples (RANDALL1). These derivation and validation samples were used during regression and discriminant model development described in Chapter IV.

TABLE XI

Program to Extract Data from the Master File

```
//ADDATA JOB (2807,0110),'D OSLUND, SMC 1763',CLASS=K
//*MAIN ORG=NPGVM1.2807P
//   EXEC SAS
//SAS.WORK DD SPACE=(CYL,(12,4))
//FILEIN DD  UNIT=3400-5,VOL=SER=ENLIST,
//   DISP=OLD,DSN=ENLIST.ALL.A7678
//FILEOUT DD UNIT=3330V,MSVGP=PUB4B,DISP=(NEW,CATLG,DELETE),
//          DSN=MSS.S2807.ADDATA,
//             DCB=(BLKSIZE=6400)
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

DATA FILEOUT.ADDATA;

* THIS PROGRAM EXTRACTS NEARLY ALL THE VARIABLES FROM THE
  MASTER FILE, AND WRITES OUT A FILE TO MASS STORAGE
  WHICH CONTAINS ALL THESE VARIABLES FOR ALL CASES WHICH
  HAD ANYTHING TO DO WITH THE 'AD' RATING.;

      INFILE FILEIN;
      INPUT

@  5 CENSUSRG PIB1.  @  6 CENSUSDS PIB1.  @  7 HOMEZIP  PIB3.
@ 10 PRESTATE PIB1.  @ 11 DATEDETY PIB1.  @ 12 DATEDETM PIB1.
@ 13 BIRTHYR  PIB1.  @ 14 BIRTHMTH PIB1.  @ 15 BIRTHDAY PIB1.
@ 16 ENTRYAGE PIB1.  @ 17 RECORDID PIB1.  @ 18 HYEC     PIB1.
@ 19 SEX      PIB1.  @ 20 RACE     PIB1.  @ 21 ETHNIC   PIB1.
@ 22 RACEETHN PIB1.  @ 23 MRTLDPND PIB1.  @ 24 TESTFORM PIB1.
@ 25 AFQTPCNT PIB1.  @ 26 AFQTGRPS PIB1.  @ 27 ASVABGI  PIB1.
@ 28 ASVABNO  PIB1.  @ 29 ASVABAD  PIB1.  @ 30 ASVABWK  PIB1.
@ 31 ASVABAR  PIB1.  @ 32 ASVABSP  PIB1.  @ 33 ASVABMK  PIB1.
@ 34 ASVABEI  PIB1.  @ 35 ASVABMC  PIB1.  @ 36 ASVABGS  PIB1.
@ 37 ASVABSI  PIB1.  @ 38 ASVABAI  PIB1.  @ 39 SERVACCS PIB1.
@ 40 PRIORSRV PIB1.  @ 41 PUL      PIB1.  @ 42 HES      PIB1.
@ 43 ASVABCM  PIB1.  @ 44 ASVABCA  PIB1.  @ 45 ASVABCE  PIB1.
@ 46 ASVABCC  PIB1.  @ 47 ENTRYSTA PIB1.  @ 48 HEIGHT   PIB1.
@ 49 WEIGHT   PIB1.  @ 50 SYSTOLBP PIB1.  @ 51 DIASTLBP PIB1.
@ 52 MEDFAIL1 PIB1.  @ 53 MEDFAIL2 PIB1.  @ 54 MEDFAIL3 PIB1.
@ 55 WAIVER   PIB1.  @ 56 WAIVERAL PIB1.  @ 57 EXAMSTAT PIB1.
@ 58 ENTRYYR  PIB1.  @ 61 TERMENLT PIB1.  @ 62 ENTRPAYG PIB1.
@ 59 ENTRYMTH PIB1.  @ 60 ENTRYDAY PIB1.
@ 63 ECMPCNTY PIB2.  @ 65 PROGENLT PIB5.  @ 72 AFEESSTA PIB1.
@ 73 ECNDSCPT PIB1.  @ 74 ENLSTCPT PIB1.  @ 75 YOUTHPRG PIB1.
@ 78 TAPEDATE PIB1.  @ 81 TRENLMOS PIB5.  @ 86 TAFMS1   PIB2.
@ 88 DPCC1    PIB2.  @ 90 DDCC1    PIB2.  @ 92 HYEC1    PIB1.
@ 93 PAYGRDE1 PIB1.  @ 94 SERVICE1 PIB1.  @ 95 MRTSTAT1 PIB1.
@ 96 NDPNDNT1 PIB1.  @ 97 SPNSPD1  PIB3.  @100 ISC1     PIB1.
@101 SEPRT1YR PIB1.  @102 SEPRT1MT PIB1.  @103 SEPRT1DY PIB1.
@104 BASD1YR  PIB1.  @105 BASD1MTH PIB1.  @106 BASD1DAY PIB1.
@107 ETS1YEAR PIB1.  @108 ETS1MNTH PIB1.
@109 DOLE1YR  PIB1.  @110 DOLE1MTH PIB1.
@113 PEBD1YR  PIB1.  @114 PEBD1MTH PIB1.  @115 PEBD1DAY PIB1.
@111 CHARSRV1 PIB1.  @112 ELGREUP1 PIB1.
@116 FILEFLG1 PIB2.  @118 TAFMS2   PIB2.
@120 DPCC2    PIB2.  @122 DDCC2    PIB2.  @124 HYEC2    PIB1.
@125 PAYGRDE2 PIB1.  @126 SERVICE2 PIB1.  @127 MRTSTAT2 PIB1.
@128 NDPNDNT2 PIB1.  @129 SPNSPD2  PIB3.  @132 ISC2     PIB1.
@133 SEPRT2YR PIB1.  @134 SEPRT2MT PIB1.  @135 SEPRT2DY PIB1.
@136 BASD2YR  PIB1.  @137 BASD2MTH PIB1.  @138 BASD2DAY PIB1.
@139 ETS2YEAR PIB1.  @140 ETS2MNTH PIB1.
@141 DOLE2YR  PIB1.  @142 DOLE2MTH PIB1.
@145 PEBD2YR  PIB1.  @146 PEBD2MTH PIB1.  @147 PEBD2DAY PIB1.
@143 CHARSRV2 PIB1.  @144 ELGREUP2 PIB1.
```

```
@148 FILEFLG2 PIB2.  @150 TAFMS3   PIB1.
@151 TAFMS4   PIB1.  @152 DPOC3    PIB2.  @154 DDOC3    PIB2.
@156 HYEC3    PIB1.  @157 PAYGRDE3 PIB1.  @158 SERVICE3 PIB1.
@159 MRTSTAT3 PIB1.  @160 NDPNDNT3 PIB1.  @161 SPNSPD3  PIB3.
@165 SEPRT3YR PIB1.  @166 SEPRT3MT PIB1.  @167 SEPRT3DY PIB1.
@168 BASD3YR  PIB1.  @169 BASD3MTH PIB1.  @170 BASD3DAY PIB1.
@171 ETS3YEAR PIB1.  @172 ETS3MNTH PIB1.
@173 DOLE3YR  PIB1.  @174 DOLE3MTH PIB1.
@177 PEBD3YR  PIB1.  @178 PEBD3MTH PIB1.  @179 PEBD3DAY PIB1.
@164 ISC3     PIB1.
@175 CHARSRV3 PIB1.
@176 ELGREUP3 PIB1.  @180 FILEFLG3 PIB2.
@182 FILEMTCH PIB4.  @186 DOEYRDEP PIB1.  @187 DOEMTDEP PIB1.
@188 MNTHSDEP PIB1.  @189 SPFLGML  PIB1.
@190 DCPGYR   PIB1.  @191 DCPGMNTH PIB1.
@212 GCT         2.  @214 ARI         2.  @216 MECH        2.
@218 CLER        2.  @220 AFQTS       2.  @222 PNEC      $4.
@227 CTZNSHIP  $1.
@229 PRIDEPND  $1.   @230 SECDEPND   $1.  @231 ERCL      $2.
@233 GPCURIND  $1.   @234 AUTHRATE   $4.  @240 EDPGYR    $4.
@244 SCHLCODE  $1.   @245 SCHLWVR    $1.  @246 ASTAR     $1.
@247 ISSIND    $1.   @250 PRESRATE   $4.
@254 NUMPG1    $1.   @255 PRFTAERV   $3.  @258 EXAMRATE  $4.
@262 NUMPG2    $1.   @263 EXRTAERV   $3.  @266 TOTLRAW    3.
@269 STDNAVY    2.   @272 PRCODE     $2.  @274 ALTPRCDE  $2.
@276 FINLMULT   5.   @281 FNMLTCUT    5.  @287 PRFFACTR   3.
@290 AWTFACTR   2.   @292 CHNGRATE   $1.
@296 RATEIND   $1.   @297 SPERCIND   $1.  @298 TYPENLST  $2.
@301 MCDEST    $1.   @302 NENLSTMT    1.
@303 RACS  YYMMDD6.  @309 TAS        $4.  @313 CAS       $4.
@317 ICSCODE   $1.   @318 LOSWVR     $1.  @319 SIPG      $4.
@323 TIRWVR    $1.   @324 TIR        $4.
@336 ADED  YYMMDD6.  @343 EDPG  YYMMDD6.  @349 DTIS       3.
@352 RECFORES   1.   @356 NCHANGES    3.  @384 AGE        2.
@386 NHRCGCT    2.   @388 NHRCAFQT    2.  @390 MENTLGRP  $1.
@391 EDCERTIF  $1.   @392 MOBLDSGN   $1.  @394 HYNDPNDT   2.
@396 GRP4PRCG  $2.   @398 SSDUTY     $1.  @399 REGRESRV  $1.
@400 HYRAYGRD  $1.   @401 NOTRCMD    $1.  @402 SSNCHNGE  $1.
@403 TOTRFCMO   2.   @405 TOTLDEMO    1.  @406 TOTLAWOL   1.
@407 TOTDESRT   1.   @408 TOTMLTCN    1.  @409 TOTCVLCN   1.
@412 LNGTHSRV  $4.   @416 SCREEN      2.  @418 ATTRITCD  $1.
@419 RECNTC    $1.   @420 RECENLST   $2.  @422 RECPRCGM  $1.
@423 RECPRGSC  $2.   @425 RCPGSCRT   $4.  @435 ELSTHIST  $1.
@436 NDAYSE2    4.   @440 NDAYSE3     4.  @444 NDAYSE4    4.
@449 DMDCRATE  $3.   @452 DMDCNEC    $4.  @456 DMDCUIC   $6.
@462 CONVDATE YYMMDD6.                    @468 GRADDATE YYMMDD6.
@474 TRANDATE YYMMDD6.
@480 RARNNEC $4.     @484 TRAININD $1.    @485 STACTION $1.;

LABEL
CENSUSRG=CENSUS REGION (10 CODES)
CENSUSDS=CENSUS DISTRICT (5 CODES)
HOMEZIP =HOME OF RECORD ZIP CODE
HMESTATE=HOME OF RECORD--STATE
DATEDETY=YEAR OF FINAL QUALIFYING DETERMINATION
DATEDETM=MONTH OF FINAL QUALIFYING DETERMINATION
BIRTHYR =YEAR OF BIRTH
BIRTHMTH=MONTH OF BIRTH
BIRTHDAY=DAY OF BIRTH
ENTRYAGE=AGE OF INDIVIDUAL AT TIME OF ENTRY
RECORDID=RECORD ID--EXAM SCORE, DEP, ACTIVE DUTY
HYEC    =HIGHEST YEAR OF EDUCATION
SEX     =(1) MALE, (2) FEMALE
RACE    =(1) WHITE, (2) BLACK, (3) OTHER
ETHNIC  =INDIVIDUAL'S REPORTED ETHNIC STATUS
RACEETHN=SIX RACE-ETHNIC COMBINATIONS
MRTLDPND=MARITAL STATUS/DEPENDENTS
TESTFORM=TEST FORM/ECFA,ASVAB,AFWST,AFQT,OSB...
AFQTPCNT=AFQT PERCENTILE (OR EQUIVALENT)
```

```
AFQTGRPS=AFQT GROUPS (5,4C,4B,4A,3B,3A,2,1)
ASVABGI =ASVAB APTITUDE AREA SCORE--SUBSCALE GI
ASVABNO =ASVAB APTITUDE AREA SCORE--SUBSCALE NO
ASVABAD =ASVAB APTITUDE AREA SCORE--SUBSCALE AD
ASVABWK =ASVAB APTITUDE AREA SCORE--SUBSCALE WK
ASVABAR =ASVAB APTITUDE AREA SCORE--SUBSCALE AR
ASVABSP =ASVAB APTITUDE AREA SCORE--SUBSCALE SP
ASVABMK =ASVAB APTITUDE AREA SCORE--SUBSCALE MK
ASVABEI =ASVAB APTITUDE AREA SCORE--SUBSCALE EI
ASVABMC =ASVAB APTITUDE AREA SCORE--SUBSCALE MC
ASVABGS =ASVAB APTITUDE AREA SCORE--SUBSCALE GS
ASVABSI =ASVAB APTITUDE AREA SCORE--SUBSCALE SI
ASVABAI =ASVAB APTITUDE AREA SCORE--SUBSCALE AI
SERVACCS=SERVICE OF ACCESSION (NAVY,2)
PRIORSRV=PRIOR SERVICE (NON-PRIOR SERVICE,1)
PUL     =GEN. HEALTH, UPPER & LOWER EXTREMITIES
HES     =HEARING, VISION, PSYCHIATRIC WELL-BEING
ASVABCM =ASVAB APTITUDE AREA SCORE--SUBSCALE CM
ASVABCA =ASVAB APTITUDE AREA SCORE--SUBSCALE CA
ASVABCE =ASVAB APTITUDE AREA SCORE--SUBSCALE CE
ASVABCC =ASVAB APTITUDE AREA SCORE--SUBSCALE CC
ENTRYSTA=ENTRY STATUS (1,DIRECT TO ACTIVE DUTY)
HEIGHT  =HEIGHT IN INCHES (FRACTIONS DROPPED)
WEIGHT  =WEIGHT IN POUNDS (FRACTIONS ROUNDED)
SYSTOLBP=BLOOD PRESSURE--SYSTOLIC
DIASTLBP=BLOOD PRESSURE--DIASTOLIC
MEDFAIL1=PRIMARY MEDICALLY DISQUALIFYING DEFECT
MEDFAIL2=SECONDARY MEDICALLY DISQUALIFYING DEFECT
MEDFAIL3=TERTIARY MEDICALLY DISQUALIFYING DEFECT
WAIVER  =PERMIT CODE FOR AN OTHERWISE INELIGIBLE
WAIVERAL=WAIVER APPROVAL LEVEL AND EXPLANATION
EXAMSTAT=EXAMINATION STATUS (1,FULLY QUALIFIED)
TERMENLI=TERM OF ENLISTMENT (NO. OF YEARS)
ENTRPAYG=ENTRY PAY GRADE (E00--O11)
HOMECNTY=HOME OF RECORD COUNTY--FIPS
PROGENLI=PROGRAM ENLISTED FOR--SERVICE UNIQUE
AFEESSTA=MILITARY ENTRANCE PROCESSING STATIONS
BONUSOPT=BONUS OPTION, COMBAT OR NON-COMBAT
ENLSTOPT=ENLISTMENT OPTION
YOUTHPRG=YOUTH & RESERVE TRAINING PROGRAMS
TAPEDATE=MONTH OF FILE ON WHICH RECORD SUBMITTED
TRENLMOS=OCCUP. SPECIAL./RATING CHOICE UPON ENTRY
TAFMS1  =MONTHS OF TOTL. ACTIVE FED. MILIT. SERV.
DPOC1   =D.O.D. PRIMARY OCCUPATION CODE
DDOC1   =D.O.D. DUTY OCCUPATION CODE
HYEC1   =HIGHEST YEAR OF EDUCATION
PAYGRDE1=PAY GRADE AS-OF-DATE-OF-FILE/SEPARATION
SERVICE1=SERVICE CODE (2, NAVY)
MRTSTAT1=MARITAL STATUS (1,OTHER, 2,MARRIED)
NDPNDNT1=NUMBER OF DEPENDENTS (1, NONE)
SPNSPD1 =SEPARATION PROGRAM DESIGNATOR
ISC1    =INTER-SERVICE SEPARATION CODE
SEPRT1YR=YEAR OF SEPARATION (2ND DMDC SECTION)
SEPRT1MT=MONTH OF SEPARATION (2ND DMDC SECTION)
SEPRT1DY=DAY OF SEPARATION (2ND DMDC SECTION)
BASD1YR =YEAR OF ACTIVE DUTY BASE DATE
BASD1MTH=MONTH OF ACTIVE DUTY BASE DATE
BASD1DAY=DAY OF ACTIVE DUTY BASE DATE
ETS1YEAR=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
ETS1MNTH=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
CHARSRV1=CHARACTER OF SERVICE
ELGREUP1=REENLISTMENT ELIGIBILITY
PEBD1YR =YEAR OF PAY ENTRY BASE DATE
PEBD1MTH=MONTH OF PAY ENTRY BASE DATE
PEBD1DAY=DAY OF PAY ENTRY BASE DATE
ENTRYYR =YEAR OF ENTRY TO ACTIVE/D.E.P.
ENTRYMTH=MONTH OF ENTRY TO ACTIVE/D.E.P.
ENTRYDAY=DAY OF ENTRY TO ACTIVE/D.E.P.
SEPRT1YR=YEAR OF SEPARATION (2ND DMDC SECTION)
```

```
SEPRT1MT=MONTH OF SEPARATION (2ND DMDC SECTION)
SEPRT1DY=DAY OF SEPARATION (2ND DMDC SECTION)
BASD1YR =YEAR OF ACTIVE DUTY BASE DATE
BASD1MTH=MONTH OF ACTIVE DUTY BASE DATE
BASD1DAY=DAY OF ACTIVE DUTY BASE DATE
ETS1YEAR=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
ETS1MNTH=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
PEBD1YR =YEAR OF PAY ENTRY BASE DATE
PEBD1MTH=MONTH OF PAY ENTRY BASE DATE
PEBD1DAY=DAY OF PAY ENTRY BASE DATE
FILEFLG1=FILE FLAG NO. 1
PEBD2YR =YEAR OF PAY ENTRY BASE DATE
PEBD2MTH=MONTH OF PAY ENTRY BASE DATE
PEBD2DAY=DAY OF PAY ENTRY BASE DATE
SEPRT2YR=YEAR OF SEPARATION (3RD DMDC SECTION)
SEPRT2MT=MONTH OF SEPARATION (3RD DMDC SECTION)
SEPRT2DY=DAY OF SEPARATION (3RD DMDC SECTION)
BASD2YR =YEAR OF ACTIVE DUTY BASE DATE
BASD2MTH=MONTH OF ACTIVE DUTY BASE DATE
BASD2DAY=DAY OF ACTIVE DUTY BASE DATE
ETS2YEAR=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
ETS2MNTH=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
PEBD2YR =YEAR OF PAY ENTRY BASE DATE
PEBD2MTH=MONTH OF PAY ENTRY BASE DATE
PEBD2DAY=DAY OF PAY ENTRY BASE DATE
TAFMS2  =MONTHS OF TOTL. ACTIVE FED. MILIT. SERV.
DPOC2   =D.O.D. PRIMARY OCCUPATION CODE
DDOC2   =D.O.D. DUTY OCCUPATION CODE
HYEC2   =HIGHEST YEAR OF EDUCATION
PAYGRDE2=PAY GRADE AS-OF-DATE-OF-FILE/SEPARATION
SERVICE2=SERVICE CODE (2, NAVY)
MRTSTAT2=MARITAL STATUS (1,OTHER, 2,MARRIED)
NDPNDNT2=NUMBER OF DEPENDENTS (1, NONE)
SPNSPD2 =SEPARATION PROGRAM DESIGNATOR
ISC2    =INTER-SERVICE SEPARATION CODE
CHARSRV2=CHARACTER OF SERVICE
ELGREUP2=REENLISTMENT ELIGIBILITY
FILEFLG2=FILE FLAG NO. 2
PEBD3YR =YEAR OF PAY ENTRY BASE DATE
PEBD3MTH=MONTH OF PAY ENTRY BASE DATE
PEBD3DAY=DAY OF PAY ENTRY BASE DATE
SEPRT3YR=YEAR OF SEPARATION (4TH DMDC SECTION)
SEPRT3MT=MONTH OF SEPARATION (4TH DMDC SECTION)
SEPRT3DY=DAY OF SEPARATION (4TH DMDC SECTION)
BASD3YR =YEAR OF ACTIVE DUTY BASE DATE
BASD3MTH=MONTH OF ACTIVE DUTY BASE DATE
BASD3DAY=DAY OF ACTIVE DUTY BASE DATE
ETS3YEAR=ESTIMATED YEAR OF FULFILLED ACTIVE DUTY
ETS3MNTH=ESTIMATED MONTH OF FULFILLED ACTIVE DUTY
PEBD3YR =YEAR OF PAY ENTRY BASE DATE
PEBD3MTH=MONTH OF PAY ENTRY BASE DATE
PEBD3DAY=DAY OF PAY ENTRY BASE DATE
TAFMS3  =MONTHS OF TOTL. ACTIVE FED. MILIT. SERV.
TAFMS4  =MONTHS OF TOTL. ACTIVE FED. MILIT. SERV.
DPOC3   =D.O.D. PRIMARY OCCUPATION CODE
DDOC3   =D.O.D. DUTY OCCUPATION CODE
HYEC3   =HIGHEST YEAR OF EDUCATION
PAYGRDE3=PAY GRADE AS-OF-DATE-OF-FILE/SEPARATION
SERVICE3=SERVICE CODE (2, NAVY)
MRTSTAT3=MARITAL STATUS (1,OTHER, 2,MARRIED)
NDPNDNT3=NUMBER OF DEPENDENTS (1, NONE)
SPNSPD3 =SEPARATION PROGRAM DESIGNATOR
ISC3    =INTER-SERVICE SEPARATION CODE
CHARSRV3=CHARACTER OF SERVICE
ELGREUP3=REENLISTMENT ELIGIBILITY
FILEFLG3=FILE FLAG NO. 3
FILEMTCH=4-BYTE BINARY FILE MATCH INDICATORS
DOEYRDEP=DOE YEAR INTO D.E.P.
DOEMTDEP=DOE MONTH INTO D.E.P.
```

```
MNTHSDEP=MONTHS IN D.E.P.
SPFLGMT =SPANISH FLAG MASTER/LOSS
DCPGMNTH=MONTH OF DCPG
DCPGYR  =YEAR OF DCPG
GCT     =BASIC BATTERY GCT
ARI     =BASIC BATTERY ARI
MECH    =BASIC BATTERY MECH
CLER    =BASIC BATTERY CLER
ENEC    =NAVY ENLISTED JOB CODE
CTZNSHIP=CITIZENSHIP CODE
BRCL    =BRANCH/CLASS
GROUPIND=GROUP INDICATOR
AUTHRATE=AUTHORIZED RATE (ABBR.)
EDPGYR  =EFFECTIVE DATE OF PAY GRADE
SCHLCODE=SCHOOL CODE
SCHLWVR =SCHOOL WAIVER
PRESRATE=PRESENT RATE CODE
PRRTABBV=PRESENT RATE (ABBR.)
EXAMRATE=EXAMINATION RATE CODE
EXRTABBV=EXAMINATION RATE (ABBR.)
TOTLRAW =TOTAL RAW SCORE
STDNAVY =STANDARDIZED NAVY SCORE
PRCODE  =PROCESS CODE
ALTPRCDE=ALTERNATE PROCESS CODE
FINLMULT=CANDIDATE''S FINAL MULTIPLE
FNMLTCUT=FINAL MULTIPLE CUT
PRFFACTR=PERFORMANCE FACTOR
AWIFACTR=AWI FACTOR
CHNGRATE=CHANGE OF RATE INDICATOR
NENLSTMT=NUMBER OF ENLISTMENTS
EAOS    =EXPIRATION OF ACTIVE OBLIGATED SERVICE
TAS     =TOTAL ACTIVE SERVICE
CAS     =OTHER ACTIVE SERVICE
SIPG    =SERVICE IN PAY GRADE
LOSCODE =LENGTH OF SERVICE
LOSWVR  =LENGTH OF SERVICE WAIVER
TIRWVR  =TIME IN RATE WAIVER
TIR     =TIME IN RATE
ADBD    =ACTIVE DUTY BASE DATE
EDPG    =EFFECTIVE DATE OF PAY GRADE
DTIS    =DRILL TIME IN SERVICE
NCHANGES=NUMBER OF CHANGES/ENTRIES IN NHRC FILE
AGE     =CANDIDATE''S CURRENT AGE
NHRCGCT =NHRC FILE''S GENRL. CLASSIFICATION TEST
NHRCAFQT=NHRC FILE''S ARMED FORCES QUALIFY. TEST
MENTLGRP=MENTAL GROUP CODE
EDCERTIF=EDUCATION CERTIFICATE
MOBLDSGN=MILITARY OBLIGATION DESIGNATOR
HYNDPNDT=HIGHEST NUMBER OF PRIMARY DEPENDENTS
GRP4PPCG=GROUP IV (100K) PROGRAM CODE
SSDUTY  =SEA-SHORE DUTY INDICATOR
REGRESRV=REGULAR RESERVE INDICATOR
HYPAYGRD=HIGHEST PAY GRADE
NOTRCMD =NOT RECOMMENDED FOR RE-ENLISTMENT
SSNCHNGE=SOCIAL SECURITY/NAME CHANGE
TOTPROMO=TOTAL PROMOTIONS
TOTLDEMO=TOTAL DEMOTIONS
TOTLAWOL=TOTAL UA/AWOL
TOTDESRT=TOTAL DESERTIONS
TOTMLTCN=TOTAL MILITARY CONFINEMENTS
TOTCVLCN=TOTAL CIVILIAN CONFINEMENTS
LNGTHSRV=LENGTH OF SERVICE
SCREEN  =SCREEN SCORE
ATTRITCD=ATTRITION INDICATOR
RECNTC  =RECRUIT NAVAL TRAINING COMMAND
RECENLST=RECRUIT TYPE ENLISTMENT
RECPRCGM=RECRUIT PROGRAM AT ENLISTMENT
RECPRGSC=RECRUIT PROGRAM/SCHOOL
RCPGSCRT=RECRUIT PROGRAM/SCHOOL RATE
```

```
ELSTHIST=ENLISTED HISTORY STATUS
NDAYSE2 =COMPUTED NUMBER OF DAYS TO E-2 RATING
NDAYSE3 =COMPUTED NUMBER OF DAYS TO E-3 RATING
NDAYSE4 =COMPUTED NUMBER OF DAYS TO E-4 RATING
DOLE1YR =YEAR OF LATEST RE-ENLISTMENT
DOLE1MTH=MONTH OF LATEST RE-ENLISTMENT
DOLE2YR =YEAR OF LATEST RE-ENLISTMENT
DOLE2MTH=MONTH OF LATEST RE-ENLISTMENT
DOLE3YR =YEAR OF LATEST RE-ENLISTMENT
DOLE3MTH=MONTH OF LATEST RE-ENLISTMENT
DMDCRATE=FINAL RATING AS LISTED BY D.M.D.C.
DMDCNEC =FINAL N.E.C. AS LISTED BY D.M.D.C.
DMDCUIC =FINAL U.I.C. AS LISTED BY D.M.D.C.
CONVDATE=CONVENING DATE FOR NITRAS COURSE
GRADDATE=GRADUATION DATE FOR NITRAS COURSE
TRANDATE=TRANSACTION DATE FOR NITRAS RECORD
EARNNEC =DID CANDIDATE EARN AN NEC?
TRAININD=TRAINING INDICATOR
STACTION=STUDENT ACTION CODES (PASS, 2, ETC.);

* THIS SCREEN SELECTS ONLY THOSE CASES WHICH HAD ANY
  AFFILIATION WITH THE 'AD' RATING.  THAT IS, THOSE CASES
  WHICH ARE LISTED IN THE DMDC FILE AS PRESENTLY AD'S
  (PRRTAERV) OR AS FINALLY AD'S (DMDCRATE), OR AS SIGNING
  UP FOR AD'S (RCPGSCRT), OR AS HAVING TAKEN THE AD
  RATING EXAMINATION (EXAMRATE).;

      IF DMDCRATE='AD' OR PRRTAERV='AD' OR
      RCPGSCRT='6200' OR EXAMRATE='6200';

* THIS NEXT SECTION OUTPUTS BASIC FREQUENCIES TO CHECK
  THAT THE RATING SPECIFIC DATA HAS BEEN WRITTEN ONTO
  THE FILE IN MASS STORAGE.;

PROC FREQ DATA=FILEOUT.ADDATA;
   TABLES DMDCRATE PRRTAERV RCPGSCRT EXAMRATE;
TITLE CHECKOUT FREQUENCIES FROM THE FILE ADDATA.;

/*
//
```

TABLE XII

Program to Screen the AD Data

```
//ADSCREEN JOB (2807,C110),'D OSLUND, SMC 1763',CLASS=B
//*MAIN ORG=NPGVM1.2807P
//   EXEC SAS
//SAS.WORK DD SPACE=(CYL,(12,4))
//FILEIN DD  DISP=SHR,DSN=MSS.S2807.ADDATA
//FILEOUT DD UNIT=3330V,MSVGP=PUB4A,
//   DISP=(NEW,CATLG,DELETE),DSN=MSS.S2807.ADSUBSET,
//   DCB=(BLKSIZE=6400)
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

* THIS PROGRAM REDUCES THE NUMBER OF CASES IN THE DATA
  SET BY SCREENING ON CERTAIN VARIABLES.  THE INTENT
  OF THE SCREEN IS SUMMARIZED ABOVE THE APPROPRIATE
  SAS STATEMENTS;

DATA FILEOUT.ADSUBSET;
   SET FILEIN.ADDATA;

* THE NUMBER OF VARIABLES IN THE DATA IS REDUCED TO
  REDUCE THE WORK SPACE REQUIREMENTS.;

     KEEP
     AFQTGRES AFQTPCNT AGE      ASVABAD  ASVABAI
     ASVABAR  ASVABEI  ASVABGI  ASVABGS  ASVABMC
     ASVABMK  ASVABNC  ASVABSI  ASVABSP  ASVABWK
     ATTRITCD AUTHRATE AWIFACTR BASD1DAY BASD1MTH
     BASD1YR  CHARSRV1 CHARSRV3 DDOC1    DDOC3
     DMDCNEC  DMDCRATE DOLE1MTH DOLE1YR  DPOC1
     DPOC3    EDCERTIF ELGREUP1 ELGREUP3 ENTRPAYG
     ENTRYAGE ENTRYDAY ENTRYMTH ENTRYSTA ENTRYYR
     ETHNIC   ETS1MNTH ETS1YEAR EXAMRATE EXRTABRV
     FILEFLG1 FILEFLG3 FINLMULT FNMLTCUT HYEC
     HYEC1    HYEC3    HYNDPNDT HYPAYGRD ISC1
     ISC3     LNGTHSRV MENTLGRP MOBLDSGN MRTLDPND
     MRTSTAT1 MRTSTAT3 NDAYSE2  NDAYSE3  NDAYSE4
     NDPNDNT1 NDPNDNT3 NHRCAFQT NOTRCMD  PAYGRDE1
     PAYGRDE3 PEBD1DAY PEBD1MTH PEBD1YR  PRESRATE
     PRFFACTR PRIORSRV PRRTAERV RACE     RACEETHN
     RCPGSCRT RECENLST RECORDID RECPRGSC REGRESRV
     SCREEN   SEPRT1DY SEPRT1MT SEPRT1YR SEPRT3DY
     SEPRT3DY SEPRT3MT SEPRT3YR SERVACCS SERVICE1
     SERVICE3 SEX      SPNSPD1  SPNSPD3  SSNCHNGE
     STDNAVY  TAFMS1   TAFMS3   TAFMS4   TERMENLT
     TESTFORM TOTCVLCN TOTDESRT TOTLAWCL TOTLDEMO
     TOTLRAW  TOTMLTCN TOTPRCMC TRNLMOS  WAIVER
     WAIVERAL;

* THIS SCREEN SELECTS ONLY THOSE CASES WHOSE FINAL
  DMDC RATING IS AD.;

     IF DMDCRATE EQ 'AD';

* THE FOLLOWING LINE SELECTS ONLY THOSE CASES WITH NO
  PRIOR SERVICE.  TO FURTHER REMOVE POTENTIAL PRIOR
  SERVICE CASES THOSE WHO HAVE CHANGED THEIR SOCIAL
  SECURITY NUMBER ARE ALSO REMOVED FROM THE SAMPLE.;

     IF PRIORSRV=1; IF SSNCHNGE EQ 0;
```

```
* THE FOLLOWING STATEMENTS SELECT ONLY THOSE CASES WHO
  WERE TESTED ON ASVAB FORMS 5, 6 OR 7.  ALSO THOSE
  CASES WITH PECULIARLY HIGH ASVAB SCORES ARE
  ELIMINATED FROM THE DATA SET.;

     IF ((TESTFORM GE 35) AND (TESTFORM LE 37));
     IF ASVABGI<=15; IF ASVABNC<=50; IF ASVABAD<=30;
     IF ASVABAR<=20; IF ASVABSP<=20; IF ASVABMK<=20;
     IF ASVABGS<=20; IF ASVABSI<=20; IF ASVABAI<=20;
     IF ASVABWK<=30; IF ASVABEI<=30; IF ASVABMC<=20;

* THIS SCREEN ONLY KEEPS THOSE WHO SIGNED UP FOR
  NAVY OR NAVAL RESERVE.;

     IF ((SERVACCS EQ 2) OR (SERVACCS EQ 8));

* ONLY THOSE CASES WHO WERE KNOWN TO HAVE SIGNED
  UP FOR AT LEAST FOUR YEARS ACTIVE DUTY ARE KEPT.;

     IF RECENLST EQ '11';

* THE CASES ARE SCREENED TO INCLUDE ONLY THOSE WITH
  'GOOD' OR 'BAD' INTERSERVICE SEPARATION CODES,
  'GREY' CASES ARE ELIMINATED.;

     IF ISC1=0 OR ISC1=1
               OR (ISC1 GE 60 AND ISC1 LT 90);
     IF ISC3=0 OR ISC3=1
               OR (ISC3 GE 60 AND ISC3 LT 90);

* THIS NEXT SCREEN KEEPS THOSE CASES FOR WHICH CLEAR
  'ELIGIBLE TO REENLIST' DATA IS AVAILABLE.;

     IF ELGREUP1=0 OR ELGREUP1=1 OR ELGREUP1=4 OR
     (ELGREUP1=240 AND (ELGREUP3=0 OR ELGREUP3=1));

* THESE SCREENS ELIMINATE CASES WITH IMPOSSIBLE DATA.;

     IF AFQTGRPS NE 0;     IF ENTRPAYG NE 0;
     IF ICSMNTHS LE 72;    IF LNGTHSRV NE 0603;
     IF RACE NE 0;         IF ETHNIC NE 0;
     IF TAFMS1 LE 72;      IF LNTHSRV NE 0;
     IF ENTRYAGE NE 77;    IF AFQTPCNT NE 0;
     IF RACEETHN NE 0;

/*
//
```

TABLE XIII

Program to Create New Variables

```
//ADNEWVAR JOB (2807,0110),'D OSLUND, SMC 1763',CLASS=B
//*MAIN ORG=NPGVM1.2807P
//  EXEC SAS
//SAS.WORK DD SPACE=(CYL,(12,4))
//FILEIN DD  DISP=SHR,DSN=MSS.S2807.ADSUBSET
//FILEOUT DD UNIT=3330V,MSVGP=PUB4A,
//  DISP=(NEW,CATLG,DELETE),DSN=MSS.S2807.ADAIL4,
//  DCB=(BLKSIZE=6400)
//SYSIN DD *
OPTIONS LS=80 NOCENTER ;

* THE PURPOSE OF THIS PROGRAM IS TO GENERATE NEW VAR-
  IABLES FOR USE IN THE ANALYSIS, EITHER BY RECODING
  ORIGINAL VARIABLES, OR BY CREATING NEW VARIABLES;

DATA FILEOUT.ADAIL4;
   SET FILEIN.ADSUBSET;

* THE FOLLOWING LINES CREATE DIFFERENT 'ENTRY GROUPS'.
                ENTRY       EXAM        DMDC
  (1)            YES         YES         YES
  (2)            YES         YES          NO
  (3)            YES          NO         YES
  (4)            YES          NO          NO
  (5)             NO         YES         YES
  (6)             NO         YES          NO
  (7)             NO          NO         YES;

   IF (RCPGSCRT='6200' AND EXAMRATE='6200' AND
       DMDCRATE='AD') THEN ENTRYGRP=1;
   IF (RCPGSCRT='6200' AND EXAMRATE='6200' AND
       DMDCRATE NE 'AD') THEN ENTRYGRP=2;
   IF (RCPGSCRT='6200' AND EXAMRATE NE '6200' AND
       DMDCRATE='AD') THEN ENTRYGRP=3;
   IF (RCPGSCRT='6200' AND EXAMRATE NE '6200' AND
       DMDCRATE NE 'AD') THEN ENTRYGRP=4;
   IF (RCPGSCRT NE '6200' AND EXAMRATE='6200' AND
       DMDCRATE='AD') THEN ENTRYGRP=5;
   IF (RCPGSCRT NE '6200' AND EXAMRATE='6200' AND
       DMDCRATE NE 'AD') THEN ENTRYGRP=6;
   IF (RCPGSCRT NE '6200' AND EXAMRATE NE '6200' AND
       DMDCRATE='AD') THEN ENTRYGRP=7;

* IN THIS SECTION, THE DMDC VARIABLE 'HYEC' IS CON-
  VERTED TO A CONTINUOUS VARIABLE REPRESENTING NUMBER
  OF YEARS OF EDUCATION;

   IF HYEC=1 THEN CHYEC=3.5; IF HYEC=2 THEN CHYEC=8;
   IF HYEC=3 THEN CHYEC=9;   IF HYEC=4 THEN CHYEC=10;
   IF HYEC=5 THEN CHYEC=11;  IF HYEC=6 THEN CHYEC=12;
   IF HYEC=7 THEN CHYEC=13;  IF HYEC=8 THEN CHYEC=14;
   IF HYEC=9 THEN CHYEC=15;  IF HYEC=10 THEN CHYEC=16;
   IF HYEC=11 THEN CHYEC=18; IF HYEC=12 THEN CHYEC=20;
   IF HYEC=13 THEN CHYEC=11.5;

* A NEW CATEGORICAL VARIABLE 'HSDG' IS NOW CREATED.
  A HIGH SCHOOL GRADUATE IS CODED A '1' AND A NON HIGH
  SCHOOL GRADUATE OR A G.E.D. IS CODED '0'.;

   IF ((HYEC LE 5) OR (HYEC EQ 13)) THEN HSDG=0;
   IF ((HYEC GE 6) AND (HYEC NE 13)) THEN HSDG=1;
```

```
* THIS SECTION CREATES NEW VARIABLES REPRESENTING
  STANDARDIZED ASVAB SCORES.;

IF ASVABGI=0 THEN SASVABGI=20;
IF ASVABGI=1 THEN SASVABGI=24;
IF ASVABGI=2 THEN SASVABGI=27;
IF ASVABGI=3 THEN SASVABGI=30;
IF ASVABGI=4 THEN SASVABGI=33;
IF ASVABGI=5 THEN SASVABGI=36;
IF ASVABGI=6 THEN SASVABGI=39;
IF ASVABGI=7 THEN SASVABGI=42;
IF ASVABGI=8 THEN SASVABGI=45;
IF ASVABGI=9 THEN SASVABGI=48;
IF ASVABGI=10 THEN SASVABGI=51;
IF ASVABGI=11 THEN SASVABGI=54;
IF ASVABGI=12 THEN SASVABGI=57;
IF ASVABGI=13 THEN SASVABGI=60;
IF ASVABGI=14 THEN SASVABGI=63;
IF ASVABGI=15 THEN SASVABGI=66;
IF ASVABAR=0 THEN SASVABAR=23;
IF ASVABAR=1 THEN SASVABAR=25;
IF ASVABAR=2 THEN SASVABAR=27;
IF ASVABAR=3 THEN SASVABAR=29;
IF ASVABAR=4 THEN SASVABAR=32;
IF ASVABAR=5 THEN SASVABAR=34;
IF ASVABAR=6 THEN SASVABAR=36;
IF ASVABAR=7 THEN SASVABAR=38;
IF ASVABAR=8 THEN SASVABAR=40;
IF ASVABAR=9 THEN SASVABAR=42;
IF ASVABAR=10 THEN SASVABAR=44;
IF ASVABAR=11 THEN SASVABAR=46;
IF ASVABAR=12 THEN SASVABAR=48;
IF ASVABAR=13 THEN SASVABAR=51;
IF ASVABAR=14 THEN SASVABAR=53;
IF ASVABAR=15 THEN SASVABAR=55;
IF ASVABAR=16 THEN SASVABAR=57;
IF ASVABAR=17 THEN SASVABAR=59;
IF ASVABAR=18 THEN SASVABAR=61;
IF ASVABAR=19 THEN SASVABAR=63;
IF ASVABAR=20 THEN SASVABAR=65;
IF ASVABSP=0 THEN SASVABSP=20;
IF ASVABSP=1 THEN SASVABSP=21;
IF ASVABSP=2 THEN SASVABSP=24;
IF ASVABSP=3 THEN SASVABSP=26;
IF ASVABSP=4 THEN SASVABSP=28;
IF ASVABSP=5 THEN SASVABSP=31;
IF ASVABSP=6 THEN SASVABSP=33;
IF ASVABSP=7 THEN SASVABSP=35;
IF ASVABSP=8 THEN SASVABSP=38;
IF ASVABSP=9 THEN SASVABSP=40;
IF ASVABSP=10 THEN SASVABSP=42;
IF ASVABSP=11 THEN SASVABSP=45;
IF ASVABSP=12 THEN SASVABSP=47;
IF ASVABSP=13 THEN SASVABSP=50;
IF ASVABSP=14 THEN SASVABSP=52;
IF ASVABSP=15 THEN SASVABSP=54;
IF ASVABSP=16 THEN SASVABSP=57;
IF ASVABSP=17 THEN SASVABSP=59;
IF ASVABSP=18 THEN SASVABSP=61;
IF ASVABSP=19 THEN SASVABSP=64;
IF ASVABSP=20 THEN SASVABSP=66;
IF ASVABMK=0 THEN SASVABMK=26;
IF ASVABMK=1 THEN SASVABMK=28;
IF ASVABGS=20 THEN SASVABGS=70;
IF ASVABAI=0 THEN SASVABAI=26;
IF ASVABAI=1 THEN SASVABAI=28;
IF ASVABAI=2 THEN SASVABAI=30;
IF ASVABAI=3 THEN SASVABAI=32;
IF ASVABAI=4 THEN SASVABAI=34;
IF ASVABMK=2 THEN SASVABMK=30;
IF ASVABMK=3 THEN SASVABMK=32;
IF ASVABMK=4 THEN SASVABMK=35;
IF ASVABMK=5 THEN SASVABMK=37;
IF ASVABMK=6 THEN SASVABMK=39;
IF ASVABMK=7 THEN SASVABMK=41;
IF ASVABMK=8 THEN SASVABMK=43;
IF ASVABMK=9 THEN SASVABMK=45;
IF ASVABMK=10 THEN SASVABMK=47;
IF ASVABMK=11 THEN SASVABMK=49;
IF ASVABMK=12 THEN SASVABMK=51;
IF ASVABMK=13 THEN SASVABMK=53;
IF ASVABMK=14 THEN SASVABMK=55;
IF ASVABMK=15 THEN SASVABMK=57;
IF ASVABMK=16 THEN SASVABMK=59;
IF ASVABMK=17 THEN SASVABMK=61;
IF ASVABMK=18 THEN SASVABMK=63;
IF ASVABMK=19 THEN SASVABMK=65;
IF ASVABMK=20 THEN SASVABMK=67;
IF ASVABMC=0 THEN SASVABMC=25;
IF ASVABMC=1 THEN SASVABMC=27;
IF ASVABMC=2 THEN SASVABMC=30;
IF ASVABMC=3 THEN SASVABMC=32;
IF ASVABMC=4 THEN SASVABMC=34;
IF ASVABMC=5 THEN SASVABMC=37;
IF ASVABMC=6 THEN SASVABMC=39;
IF ASVABMC=7 THEN SASVABMC=41;
IF ASVABMC=8 THEN SASVABMC=43;
IF ASVABMC=9 THEN SASVABMC=46;
IF ASVABMC=10 THEN SASVABMC=48;
IF ASVABMC=11 THEN SASVABMC=50;
IF ASVABMC=12 THEN SASVABMC=53;
IF ASVABMC=13 THEN SASVABMC=55;
IF ASVABMC=14 THEN SASVABMC=57;
IF ASVABMC=15 THEN SASVABMC=60;
IF ASVABMC=16 THEN SASVABMC=62;
IF ASVABMC=17 THEN SASVABMC=64;
IF ASVABMC=18 THEN SASVABMC=66;
IF ASVABMC=19 THEN SASVABMC=69;
IF ASVABMC=20 THEN SASVABMC=71;
IF ASVABGS=0 THEN SASVABGS=24;
IF ASVABGS=1 THEN SASVABGS=26;
IF ASVABGS=2 THEN SASVABGS=29;
IF ASVABGS=3 THEN SASVABGS=31;
IF ASVABGS=4 THEN SASVABGS=33;
IF ASVABGS=5 THEN SASVABGS=36;
IF ASVABGS=6 THEN SASVABGS=38;
IF ASVABGS=7 THEN SASVABGS=40;
IF ASVABGS=8 THEN SASVABGS=42;
IF ASVABGS=9 THEN SASVABGS=45;
IF ASVABGS=10 THEN SASVABGS=47;
IF ASVABGS=11 THEN SASVABGS=49;
IF ASVABGS=12 THEN SASVABGS=52;
IF ASVABGS=13 THEN SASVABGS=54;
IF ASVABGS=14 THEN SASVABGS=56;
IF ASVABGS=15 THEN SASVABGS=58;
IF ASVABGS=16 THEN SASVABGS=61;
IF ASVABGS=17 THEN SASVABGS=63;
IF ASVABGS=18 THEN SASVABGS=65;
IF ASVABGS=19 THEN SASVABGS=68;
IF ASVABWK=30 THEN SASVABWK=64;
IF ASVABAD=0 THEN SASVABAD=20;
IF ASVABAD=1 THEN SASVABAD=20;
IF ASVABAD=2 THEN SASVABAD=20;
IF ASVABAD=3 THEN SASVABAD=21;
IF ASVABAD=4 THEN SASVABAD=24;
```

```
IF ASVABAI=5 THEN SASVABAI=36;
IF ASVABAI=6 THEN SASVABAI=38;
IF ASVABAI=7 THEN SASVABAI=40;
IF ASVABAI=8 THEN SASVABAI=42;
IF ASVABAI=9 THEN SASVABAI=44;
IF ASVABAI=10 THEN SASVABAI=46;
IF ASVABAI=11 THEN SASVABAI=48;
IF ASVABAI=12 THEN SASVABAI=50;
IF ASVABAI=13 THEN SASVABAI=52;
IF ASVABAI=14 THEN SASVABAI=55;
IF ASVABAI=15 THEN SASVABAI=57;
IF ASVABAI=16 THEN SASVABAI=59;
IF ASVABAI=17 THEN SASVABAI=61;
IF ASVABAI=18 THEN SASVABAI=63;
IF ASVABAI=19 THEN SASVABAI=65;
IF ASVABAI=20 THEN SASVABAI=67;
IF ASVABSI=0 THEN SASVABSI=20;
IF ASVABSI=1 THEN SASVABSI=21;
IF ASVABSI=2 THEN SASVABSI=23;
IF ASVABSI=3 THEN SASVABSI=25;
IF ASVABSI=4 THEN SASVABSI=28;
IF ASVABSI=5 THEN SASVABSI=30;
IF ASVABSI=6 THEN SASVABSI=32;
IF ASVABSI=7 THEN SASVABSI=35;
IF ASVABSI=8 THEN SASVABSI=37;
IF ASVABSI=9 THEN SASVABSI=39;
IF ASVABSI=10 THEN SASVABSI=42;
IF ASVABSI=11 THEN SASVABSI=44;
IF ASVABSI=12 THEN SASVABSI=46;
IF ASVABSI=13 THEN SASVABSI=48;
IF ASVABSI=14 THEN SASVABSI=51;
IF ASVABSI=15 THEN SASVABSI=53;
IF ASVABSI=16 THEN SASVABSI=55;
IF ASVABSI=17 THEN SASVABSI=58;
IF ASVABSI=18 THEN SASVABSI=60;
IF ASVABSI=19 THEN SASVABSI=62;
IF ASVABSI=20 THEN SASVABSI=65;
IF ASVABWK=0 THEN SASVABWK=23;
IF ASVABWK=1 THEN SASVABWK=24;
IF ASVABWK=2 THEN SASVABWK=26;
IF ASVABWK=3 THEN SASVABWK=27;
IF ASVABWK=4 THEN SASVABWK=28;
IF ASVABWK=5 THEN SASVABWK=30;
IF ASVABWK=6 THEN SASVABWK=31;
IF ASVABWK=7 THEN SASVABWK=33;
IF ASVABWK=8 THEN SASVABWK=34;
IF ASVABWK=9 THEN SASVABWK=35;
IF ASVABWK=10 THEN SASVABWK=37;
IF ASVABWK=11 THEN SASVABWK=38;
IF ASVABWK=12 THEN SASVABWK=39;
IF ASVABWK=13 THEN SASVABWK=41;
IF ASVABWK=14 THEN SASVABWK=42;
IF ASVABWK=15 THEN SASVABWK=44;
IF ASVABWK=16 THEN SASVABWK=45;
IF ASVABWK=17 THEN SASVABWK=46;
IF ASVABWK=18 THEN SASVABWK=48;
IF ASVABWK=19 THEN SASVABWK=49;
IF ASVABWK=20 THEN SASVABWK=50;
IF ASVABWK=21 THEN SASVABWK=52;
IF ASVABWK=22 THEN SASVABWK=53;
IF ASVABWK=23 THEN SASVABWK=55;
IF ASVABWK=24 THEN SASVABWK=56;
IF ASVABWK=25 THEN SASVABWK=57;
IF ASVABWK=26 THEN SASVABWK=59;
IF ASVABWK=27 THEN SASVABWK=60;
IF ASVABWK=28 THEN SASVABWK=62;
IF ASVABWK=29 THEN SASVABWK=63;
IF ASVABNO=10 THEN SASVABNO=29;
IF ASVABNO=11 THEN SASVABNO=30;
IF ASVABAD=5 THEN SASVABAD=26;
IF ASVABAD=6 THEN SASVABAD=29;
IF ASVABAD=7 THEN SASVABAD=31;
IF ASVABAD=8 THEN SASVABAD=34;
IF ASVABAD=9 THEN SASVABAD=36;
IF ASVABAD=10 THEN SASVABAD=39;
IF ASVABAD=11 THEN SASVABAD=41;
IF ASVABAD=12 THEN SASVABAD=44;
IF ASVABAD=13 THEN SASVABAD=46;
IF ASVABAD=14 THEN SASVABAD=49;
IF ASVABAD=15 THEN SASVABAD=51;
IF ASVABAD=16 THEN SASVABAD=54;
IF ASVABAD=17 THEN SASVABAD=57;
IF ASVABAD=18 THEN SASVABAD=59;
IF ASVABAD=19 THEN SASVABAD=62;
IF ASVABAD=20 THEN SASVABAD=64;
IF ASVABAD=21 THEN SASVABAD=67;
IF ASVABAD=22 THEN SASVABAD=69;
IF ASVABAD=23 THEN SASVABAD=72;
IF ASVABAD=24 THEN SASVABAD=74;
IF ASVABAD=25 THEN SASVABAD=77;
IF ASVABAD=26 THEN SASVABAD=79;
IF ASVABAD=27 THEN SASVABAD=80;
IF ASVABAD=28 THEN SASVABAD=80;
IF ASVABAD=29 THEN SASVABAD=80;
IF ASVABAD=30 THEN SASVABAD=80;
IF ASVABEI=0 THEN SASVABEI=20;
IF ASVABEI=1 THEN SASVABEI=20;
IF ASVABEI=2 THEN SASVABEI=21;
IF ASVABEI=3 THEN SASVABEI=22;
IF ASVABEI=4 THEN SASVABEI=24;
IF ASVABEI=5 THEN SASVABEI=26;
IF ASVABEI=6 THEN SASVABEI=27;
IF ASVABEI=7 THEN SASVABEI=29;
IF ASVABEI=8 THEN SASVABEI=31;
IF ASVABEI=9 THEN SASVABEI=32;
IF ASVABEI=10 THEN SASVABEI=34;
IF ASVABEI=11 THEN SASVABEI=36;
IF ASVABEI=12 THEN SASVABEI=37;
IF ASVABEI=13 THEN SASVABEI=39;
IF ASVABEI=14 THEN SASVABEI=41;
IF ASVABEI=15 THEN SASVABEI=42;
IF ASVABEI=16 THEN SASVABEI=44;
IF ASVABEI=17 THEN SASVABEI=46;
IF ASVABEI=18 THEN SASVABEI=48;
IF ASVABEI=19 THEN SASVABEI=49;
IF ASVABEI=20 THEN SASVABEI=51;
IF ASVABEI=21 THEN SASVABEI=53;
IF ASVABEI=22 THEN SASVABEI=54;
IF ASVABEI=23 THEN SASVABEI=56;
IF ASVABEI=24 THEN SASVABEI=58;
IF ASVABEI=25 THEN SASVABEI=59;
IF ASVABEI=26 THEN SASVABEI=61;
IF ASVABEI=27 THEN SASVABEI=63;
IF ASVABEI=28 THEN SASVABEI=64;
IF ASVABEI=29 THEN SASVABEI=66;
IF ASVABEI=30 THEN SASVABEI=68;
IF ASVABNO=0 THEN SASVABNO=20;
IF ASVABNO=1 THEN SASVABNO=20;
IF ASVABNO=2 THEN SASVABNO=21;
IF ASVABNO=3 THEN SASVABNO=22;
IF ASVABNO=4 THEN SASVABNO=23;
IF ASVABNO=5 THEN SASVABNO=24;
IF ASVABNO=6 THEN SASVABNO=25;
IF ASVABNO=7 THEN SASVABNO=26;
IF ASVABNO=8 THEN SASVABNO=27;
IF ASVABNO=9 THEN SASVABNO=28;
IF ASVABNO=31 THEN SASVABNO=50;
IF ASVABNO=32 THEN SASVABNO=51;
```

```
IF ASVABNO=12 THEN SASVABNO=31; IF ASVABNO=33 THEN SASVABNO=52;
IF ASVABNO=13 THEN SASVABNO=32; IF ASVABNO=34 THEN SASVABNO=53;
IF ASVABNO=14 THEN SASVABNO=33; IF ASVABNO=35 THEN SASVABNO=54;
IF ASVABNO=15 THEN SASVABNO=34; IF ASVABNO=36 THEN SASVABNO=55;
IF ASVABNO=16 THEN SASVABNO=35; IF ASVABNO=37 THEN SASVABNO=56;
IF ASVABNO=17 THEN SASVABNO=36; IF ASVABNO=38 THEN SASVABNO=57;
IF ASVABNO=18 THEN SASVABNO=37; IF ASVABNO=39 THEN SASVABNO=58;
IF ASVABNO=19 THEN SASVABNO=38; IF ASVABNO=40 THEN SASVABNO=59;
IF ASVABNO=20 THEN SASVABNO=39; IF ASVABNO=41 THEN SASVABNO=60;
IF ASVABNO=21 THEN SASVABNO=40; IF ASVABNO=42 THEN SASVABNO=61;
IF ASVABNO=22 THEN SASVABNO=41; IF ASVABNO=43 THEN SASVABNO=62;
IF ASVABNO=23 THEN SASVABNO=42; IF ASVABNO=44 THEN SASVABNO=63;
IF ASVABNO=24 THEN SASVABNO=43; IF ASVABNO=45 THEN SASVABNO=64;
IF ASVABNO=25 THEN SASVABNO=44; IF ASVABNO=46 THEN SASVABNO=65;
IF ASVABNO=26 THEN SASVABNO=45; IF ASVABNO=47 THEN SASVABNO=66;
IF ASVABNO=27 THEN SASVABNO=46; IF ASVABNO=48 THEN SASVABNO=67;
IF ASVABNO=28 THEN SASVABNO=47; IF ASVABNO=49 THEN SASVABNO=68;
IF ASVABNO=29 THEN SASVABNO=48; IF ASVABNO=50 THEN SASVABNO=69;
IF ASVABNO=30 THEN SASVABNO=49;

* THE FOLLOWING STATEMENTS CREATE THE NUMERIC VARIABLE
  'LOSMNTHS' FROM THE VARIABLE 'LNTHSRV'.;

   YEAR=SUBSTR(LNGTHSRV,1,2);
   MONTH=SUBSTR(LNGTHSRV,3,2);
   YEARS=YEAR+0;
   MONTHS=MONTH+0;
   LOSMNTHS=YEARS*12+MONTHS;

* RECODING TO A CATEGORICAL VARIABLE.;

   IF MBILDEND=10 THEN DEPENDTS=0; ELSE DEPENDTS=1;

* CONVERTING CHARACTER VARIABLES TO NUMERIC.;

   NUBYPAY=BYPAYGRD+0;   NUNOTRC=NOTRCMD+0;

* TO DEFINE THE HIGHEST PAYGRADE ACHIEVED, ACCORDING
  TO THE DMDC FILE.;

   IF FILEFIG1=8209 THEN PAYGRADE=PAYGRDE1;
   IF FILEFIG1 NE 8209 THEN PAYGRADE=PAYGRDE3;
   IF PAYGRADE=0 THEN PAYGRADE=PAYGRDE1;
   IF PAYGRADE=0 THEN PAYGRADE='.';

* CREATING THE ASVAB COMPOSITE VARIABLE USED WHEN
  CLASSIFYING AD'S, AND ASSIGNING A DUMMY VARIABLE
  TO IDENTIFY THOSE WHO ACHIEVED THE MINIMUM SCORE.;

   ADCOMPCS = SASVABAR+SASVABEI+SASVABGS+SASVABMK;
   IF ADCOMPCS GE 190 THEN ADMINSCR=1;
      ELSE ADMINSCR=0;

* SETTING UP DUMMY VARIABLES TO ALLOW ANALYSIS OF
  RACE AND SEX EFFECTS.;

   IF RACE=1 THEN WHITE=1;     ELSE WHITE=0;
   IF RACE=2 THEN BLACK=1;     ELSE BLACK=0;
   IF RACE=3 THEN OTHER=1;     ELSE OTHER=0;
   IF SEX =2 THEN NUSEX=0;     ELSE NUSEX=1;

* CREATING A RANDOM VARIABLE TO ALLOW THE DATA TO
  BE SPLIT RANDOMLY IN HALF;

   IF RANUNI(0) <= .5 THEN RANDALL1=1;  ELSE RANDALL1=0;
```

```
* CREATING INTERACTION VARIABLES FOR USE IN THE MODEL
  DEVELOPMENT.;

   INTERC1=DEPENDTS*HSDG;
   INTERC2=DEPENDTS*BLACK;
   INTERC3=DEPENDTS*NUSEX;
   INTERC4=DEPENDTS*TERMENLT;
   INTERC5=DEPENDTS*SASVABAI;
   INTERC6=DEPENDTS*ADMINSCR;
   INTERC7=HSDG*BLACK;
   INTERC8=HSDG*NUSEX;
   INTERC9=HSDG*TERMENLT;
   INTER10=HSDG*SASVABAI;
   INTER11=HSDG*ADMINSCR;
   INTER12=BLACK*NUSEX;
   INTER13=BLACK*TERMENLT;
   INTER14=BLACK*SASVABAI;
   INTER15=BLACK*ADMINSCR;
   INTER16=NUSEX*TERMENLT;
   INTER17=NUSEX*SASVABAI;
   INTER18=NUSEX*ADMINSCR;
   INTER19=TERMENLT*SASVABAI;
   INTER20=TERMENLT*ADMINSCR;
   INTER21=SASVABAI*ADMINSCR;

* THE FOLLOWING LINES CREATE DIFFERENT CRITERION
  VARIABLES.;

   IF ((SERVICE1 EQ 2) AND ((PAYGRADE GE 4) AND
        (NUHYPAY GE 4))) THEN SUCCPAYG=1;
        ELSE SUCCPAYG=0;
   IF ENTRYYR=78 AND ENTRYMTH GE 10 THEN LATEENLT=1;
        ELSE LATEENLT=0;
   IF TAFMS1 GE 48 OR (TAFMS1 GE 45 AND LATEENLT=1)
          THEN SUCCTAF=1; ELSE SUCCTAF=0;
   IF ELGREUP1=4 THEN SUCCREUP=0; ELSE SUCCREUP=1;
   IF SUCCREUP=1 AND SUCCTAF=1 AND SUCCPAYG=1
          THEN SUCCESS2=1; ELSE SUCCESS2=0;

LABEL
HSDG     =HIGH SCHOOL GRADUATE(1), OTHER(0)
DEPENDTS=SINGLE, NO DEPENDENTS (0), OTHERWISE (1)
CHYEC    =CONVERTED NUMBER OF YEARS OF EDUCATION
NUHYPAY =NHRC FILE--HIGHEST PAYGRADE ATTAINED
NUNOTRC =NHRC--NOT RECOMMENDED FOR RE-ENLISTMENT
PAYGRADE=DMDC-BASED HIGHEST PAY-GRADE ATTAINED
SASVABGI=STANDARDIZED SCORE - GENERAL INFORMATION
SASVABNO=STANDARDIZED SCORE -NUMERICAL OPERATIONS
SASVABAD=STANDARDIZED SCORE - ATTENTION TO DETAIL
SASVABWK=STANDARDIZED SCORE - WORD KNOWLEDGE
SASVABAR=STANDARDIZED SCORE -ARITHMETIC REASONING
SASVABSP=STANDARDIZED SCORE - SPACE PERCEPTION
SASVABMK=STANDARDIZED SCORE - MATH KNOWLEDGE
SASVABEI=STANDARDIZED SCORE - ELECTRONIC INFO
SASVABMC=STANDARDIZED SCORE - MECH COMPREHENSION
SASVABGS=STANDARDIZED SCORE - GENERAL SCIENCE
SASVABSI=STANDARDIZED SCORE - SHOP INFORMATION
SASVABAI=STANDARDIZED SCORE - AUTO INFORMATION
WHITE    =(1) WHITE, ELSE (0)
BLACK    =(1) BLACK, ELSE (0)
OTHER    =(1) NEITHER BLACK NOR WHITE, ELSE (0)
NUSEX    =(1) MALE, (0) FEMALE
ADCOMPCS=AD ASVAB COMPOSITE
ADMINSCR=AD ASVAB COMPOSITE SCREEN
RANDALL1=VAR. TO ALLOW A RANDOM 50-50 SPLIT
LOSMNTHS=LENGTH OF SERVICE IN MONTHS
ENTRYGRP=ENTRY GROUP CLASSIFICATIONS
LATEENLT=ENTERED AFTER SEP 78 (1), OTHERWISE (0)
SUCCTAF =SUCCESS ON LOS CRITERION (1)
```

```
SUCCPAYG=(1,0) SUCCESS ON PAYGRADE
SUCCREUP=(1,0) ELIGIBLE TO REENLIST
SUCCESS2=SUCCESS ON COMPOSITE CRITERION (1)
INTERC1=DEPENDTS*HSDG
INTERC2=DEPENDTS*BLACK
INTERC3=DEPENDTS*NUSEX
INTERC4=DEPENDTS*TERMENLT
INTERC5=DEPENDTS*SASVABAI
INTERC6=DEPENDTS*ADMINSCR
INTERC7=HSDG*BLACK
INTERC8=HSDG*NUSEX
INTERC9=HSDG*TERMENLT
INTER10=HSDG*SASVABAI
INTER11=HSDG*ADMINSCR
INTER12=BLACK*NUSEX
INTER13=BLACK*TERMENLT
INTER14=BLACK*SASVABAI
INTER15=BLACK*ADMINSCR
INTER16=NUSEX*TERMENLT
INTER17=NUSEX*SASVABAI
INTER18=NUSEX*ADMINSCR
INTER19=TERMENLT*SASVABAI
INTER20=TERMENLT*ADMINSCR
INTER21=SASVABAI*ADMINSCR;

/*
//
```

APPENDIX B

DESCRIPTIVE ANALYSIS RESULTS

Frequency distributions and correlations used for descriptive analysis of the AD data set are contained in Tables XIV and XV.

The frequencies show that 92 percent of the AD data set were 17 to 21 years of age, 79 percent had a high school degree, 97 percent were single, and 98 percent were male. Even though BLACK and OTHER only represented 17 and 6 percent of the sample respectively, their criterion scores were significantly different compared to WHITE criterion scores. Thus, BLACK and OTHER emerged as predictors in some of the models. It is interesting to note that 40 percent of the sample achieved the paygrade E-5. Using achievement of E-5 rather than E-4 in the composite success criterion would produce greater variability on the criterion which may improve the models.

One third of the cases in the data did not score 190 or greater on the AD composite score. These cases are either people who were classified prior to correcting the ASVAB Forms 5,6 and 7 misnorming problems, or people who migrated to the AD rating subsequent to service entry. This may partially explain the negative correlations these variables have with the criteria.

TABLE XIV

Selected Frequencies

FINAL RATING AS LISTED BY D.M.D.C.

DECGRADE	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
AL	2820	2820	100.000	100.000

SCREEN SCORE

SCREEN	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
.	73	.	.	.
52	2	2	0.073	0.073
55	1	3	0.036	0.109
57	5	8	0.182	0.291
59	3	11	0.109	0.400
61	1	12	0.036	0.437
62	6	18	0.218	0.655
63	7	25	0.255	0.910
64	3	28	0.109	1.019
66	36	64	1.311	2.330
68	54	118	1.966	4.296
70	66	184	2.403	6.698
71	24	208	0.874	7.572
72	145	353	5.278	12.850
73	5	358	0.182	13.032
74	106	464	3.859	16.891
75	50	514	1.820	18.711
76	37	551	1.347	20.058
77	78	629	2.839	22.898
78	188	817	6.844	29.742
79	172	989	6.261	36.003
80	131	1120	4.769	40.772
81	105	1225	3.822	44.594
82	51	1276	1.857	46.451
83	569	1845	20.714	67.164
84	90	1935	3.276	70.440
85	2	1937	0.073	70.513
86	13	1950	0.473	70.987
87	111	2061	4.041	75.027
88	424	2485	15.435	90.462
89	27	2512	0.983	91.445
90	208	2720	7.572	99.017
91	3	2723	0.109	99.126
92	5	2728	0.182	99.308
93	4	2732	0.146	99.454
94	1	2733	0.036	99.490
95	14	2747	0.510	100.000

AFQT GROUPS (5,4C,4B,4A,3B,3A,2,1)

AFQTGRPS	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	4	4	0.142	0.142
2	61	65	2.163	2.305
3	280	345	9.929	12.234
4	599	944	21.241	33.475
5	795	1739	28.191	61.667
6	545	2284	19.326	80.993
7	505	2789	17.908	98.901
8	31	2820	1.099	100.000

AGE OF INDIVIDUAL AT TIME OF ENTRY

ENTRYAGE	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
17	388	388	13.759	13.759
18	1225	1613	43.440	57.199
19	594	2207	21.064	78.262
20	263	2470	9.326	87.589
21	136	2606	4.823	92.411
22	75	2681	2.660	95.071
23	49	2730	1.738	96.809
24	28	2758	0.993	97.801
25	22	2780	0.780	98.582
26	11	2791	0.390	98.972
27	12	2803	0.426	99.397
28	8	2811	0.284	99.681
29	8	2819	0.284	99.965
30	1	2820	0.035	100.000

ENTRY PAY GRADE (E00--011)

ENTRPAYG	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	2375	2375	84.220	84.220
2	279	2654	9.894	94.113
3	166	2820	5.887	100.000

TERM OF ENLISTMENT (NO. OF YEARS)

TERMENLT	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
2	1	1	0.035	0.035
3	1	2	0.035	0.071
4	2692	2694	95.461	95.532
5	1	2695	0.035	95.567
6	125	2820	4.433	100.000

SERVICE OF ACCESSION (NAVY,2)

SERVACCS	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
2	2715	2715	96.277	96.277
8	105	2820	3.723	100.000

CONVERTED NUMBER OF YEARS OF EDUCATION

CEYEC	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
3.5	1	1	0.035	0.035
8	4	5	0.142	0.177
9	27	32	0.957	1.135
10	143	175	5.071	6.206
11	285	460	10.106	16.312
11.5	122	582	4.326	20.638
12	2165	2747	76.773	97.411
13	29	2776	1.028	98.440
14	26	2802	0.922	99.362
15	7	2809	0.248	99.610
16	11	2820	0.390	100.000

HIGH-SCHOOL GRADUATE(1) V. OTHER(0)

HSDG	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	582	582	20.638	20.638
1	2238	2820	79.362	100.000

SINGLE, NO DEPENDENTS (0), OTHERWISE (1)

DEPENDTS	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	2738	2738	97.092	97.092
1	82	2820	2.908	100.000

(1) MALE, (0) FEMALE.

NCSEX	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	64	64	2.270	2.270
1	2756	2820	97.730	100.000

ENTRY GROUP CLASSIFICATIONS

ENTRYGRP	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	1166	1166	41.348	41.348
3	128	1294	4.539	45.887
5	1316	2610	46.667	92.553
7	210	2820	7.447	100.000

(1) WHITE, (2) BLACK, (3) OTHER

RACE	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	2184	2184	77.447	77.447
2	468	2652	16.596	94.043
3	168	2820	5.957	100.000

AD ASVAB COMPOSITE SCREEN

ADMINSCR	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	945	945	33.511	33.511
1	1875	2820	66.489	100.000

VAR TO ALLOW A RANDOM 50-50 SPLIT

RANDALL1	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1380	1380	48.936	48.936
1	1440	2820	51.064	100.000

INTER-SERVICE SEPARATION CODE

ISC3	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	1106	1106	39.220	39.220
1	1495	2601	53.014	92.234
60	22	2623	0.780	93.014
61	6	2629	0.213	93.227
63	1	2630	0.035	93.262
64	7	2637	0.248	93.511
65	61	2698	2.163	95.674
67	14	2712	0.496	96.170
71	7	2719	0.248	96.418
73	15	2734	0.532	96.950
74	1	2735	0.035	96.986
75	2	2737	0.071	97.057
76	7	2744	0.248	97.305
78	22	2766	0.780	98.085
80	4	2770	0.142	98.227
82	20	2790	0.709	98.936
86	30	2820	1.064	100.000

(1) ENTERED AFTER SEP 78, OTHERWISE (0)

LATEENLT	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	2543	2543	90.177	90.177
1	277	2820	9.823	100.000

DMDC-BASED HIGHEST PAY-GRADE ATTAINED

PAYGRADE	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	110	110	3.901	3.901
2	108	218	3.830	7.730
3	231	449	8.191	15.922
4	1259	1708	44.645	60.567
5	1110	2818	39.362	99.929
6	2	2820	0.071	100.000

NHRC FILE--HIGHEST PAYGRADE ATTAINED

NUHYPAY	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
1	17	17	0.603	0.603
2	95	112	3.369	3.972
3	220	332	7.801	11.773
4	1401	1733	49.681	61.454
5	1087	2820	38.546	100.000

SUCCESS ON LCS CRITERION (1)

SUCCTAF	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	393	393	13.936	13.936
1	2427	2820	86.064	100.000

HIGH PAYGRADE SUCCESS CRITERION.

SUCCPAYG	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	474	474	16.809	16.809
1	2346	2820	83.191	100.000

REENLISTMENT ELIGIBILITY CRITERION.

SUCCREUP	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	293	293	10.390	10.390
1	2527	2820	89.610	100.000

SUCCESS ON COMPOSITE CRITERION (1)

SUCCESS2	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
0	655	655	23.227	23.227
1	2165	2820	76.773	100.000

MONTHS OF TOTL. ACTIVE FED. MILIT. SERV.

TAFMS1	FREQUENCY	CUM FREQ	PERCENT	CUM PERCENT
2	1	1	0.035	0.035
7	4	5	0.142	0.177
8	2	7	0.071	0.248
9	6	13	0.213	0.461
10	1	14	0.035	0.496
11	3	17	0.106	0.603
12	8	25	0.284	0.887
13	9	34	0.319	1.206
14	5	39	0.177	1.383
15	5	44	0.177	1.560

16	9	53	0.319	1.879
17	5	58	0.177	2.057
18	9	67	0.319	2.376
19	11	78	0.390	2.766
20	3	81	0.106	2.872
21	5	86	0.177	3.050
22	5	91	0.177	3.227
23	10	101	0.355	3.582
24	9	110	0.319	3.901
25	5	115	0.177	4.078
26	10	125	0.355	4.433
27	8	133	0.284	4.716
28	7	140	0.248	4.965
29	5	145	0.177	5.142
30	2	147	0.071	5.213
31	7	154	0.248	5.461
32	4	158	0.142	5.603
33	5	163	0.177	5.780
34	4	167	0.142	5.922
35	3	170	0.106	6.028
36	7	177	0.248	6.277
37	10	187	0.355	6.631
38	7	194	0.248	6.879
39	5	199	0.177	7.057
40	7	206	0.248	7.305
41	6	212	0.213	7.518
42	6	218	0.213	7.730
43	5	223	0.177	7.908
44	4	227	0.142	8.050
45	76	303	2.695	10.745
46	91	394	3.227	13.972
47	248	642	8.794	22.766
48	1113	1755	39.468	62.234
49	97	1852	3.440	65.674
50	91	1943	3.227	68.901
51	111	2054	3.936	72.837
52	56	2110	1.986	74.823
53	51	2161	1.809	76.631
54	61	2222	2.163	78.794
55	57	2279	2.021	80.816
56	55	2334	1.950	82.766
57	48	2382	1.702	84.468
58	34	2416	1.206	85.674
59	57	2473	2.021	87.695
60	53	2526	1.879	89.574
61	59	2585	2.092	91.667
62	53	2638	1.879	93.546
63	36	2674	1.277	94.823
64	30	2704	1.064	95.887
65	16	2720	0.567	96.454
66	22	2742	0.780	97.234
67	13	2755	0.461	97.695
68	20	2775	0.709	98.404
69	15	2790	0.532	98.936
70	11	2801	0.390	99.326
71	17	2818	0.603	99.929
72	2	2820	0.071	100.000

TABLE XV

Selected Correlations

	TAFMS1	SUCCTAF	SUCCESS2
AFQTGRPS	-0.08930 0.0001	-0.06108 0.0012	-0.05137 0.0064
AFQTPCNT	-0.07755 0.0001	-0.05346 0.0045	-0.04182 0.0264
ENTRYAGE	0.05518 0.0034	0.02593 0.1686	0.05698 0.0025
ENTRPAYG	0.03578 0.0575	-0.00926 0.6230	0.01083 0.5653
TERMENLT	0.14720 0.0001	0.02116 0.2614	0.05189 0.0059
CRYEC	0.07522 0.0001	0.07554 0.0001	0.11471 0.0001
HSDG	0.09918 0.0001	0.12117 0.0001	0.15525 0.0001
NOSEX	0.08426 0.0001	0.04868 0.0097	0.01204 0.5229
WHITE	-0.10983 0.0001	-0.06035 0.0013	-0.04767 0.0114
BLACK	0.10220 0.0001	0.05290 0.0050	0.02641 0.1608
OTHER	0.03329 0.0771	0.02342 0.2139	0.04265 0.0235
SCREEN	-0.00478 0.8022	0.07461 0.0001	0.08891 0.0001
ADCOMPCS	-0.05463 0.0037	-0.02132 0.2578	0.00440 0.8153
ADMINSCR	-0.07581 0.0001	-0.02971 0.1147	-0.02934 0.1192
SASVABAD	0.00263 0.8888	0.01025 0.5864	0.01356 0.4717
SASVABAI	-0.06941 0.0002	-0.03415 0.0698	-0.00654 0.7285
SASVABAR	-0.05163 0.0061	-0.02568 0.1727	-0.01734 0.3574
SASVABEI	-0.03140 0.0955	-0.01303 0.4893	0.01707 0.3648
SASVABGI	-0.01535 0.4152	-0.02107 0.2633	-0.00708 0.7070
SASVABMC	-0.06570 0.0005	-0.04088 0.0300	-0.02361 0.2101
SASVABMK	-0.03166 0.0928	0.00698 0.7112	0.02288 0.2245

SASVAENC	-0.03869	-0.01504	-0.00541
	0.0399	0.4246	0.7741
SASVAESI	-0.03834	-0.00090	-0.00387
	0.0418	0.9618	0.8370
SASVAESE	-0.04680	-0.02609	-0.00735
	0.0129	0.1660	0.6964
SASVAEGS	-0.03464	-0.02632	-0.01036
	0.0659	0.1622	0.5822
SASVAEWK	-0.06134	-0.05225	-0.04783
	0.0011	0.0055	0.0111
INTER01	0.05262	0.03656	0.04814
	0.0052	0.0523	0.0106
INTERC2	0.05869	0.02943	0.04022
	0.0018	0.1182	0.0327
INTERC3	0.07253	0.03665	0.04668
	0.0001	0.0516	0.0132
INTERC4	0.06297	0.03358	0.04590
	0.0008	0.0746	0.0148
INTERC5	0.06033	0.03101	0.04114
	0.0013	0.0996	0.0289
INTERC6	0.01791	0.02303	0.02299
	0.3417	0.2215	0.2222
INTERC7	0.09619	0.05711	0.03899
	0.0001	0.0024	0.0384
INTERC8	0.11832	0.12665	0.14906
	0.0001	0.0001	0.0001
INTERC9	0.12282	0.11899	0.16000
	0.0001	0.0001	0.0001
INTER10	0.07658	0.10700	0.14937
	0.0001	0.0001	0.0001
INTER11	0.00621	0.04010	0.06994
	0.7416	0.0332	0.0002
INTER12	0.10322	0.05193	0.02540
	0.0001	0.0058	0.1775
INTER13	0.10985	0.05469	0.02934
	0.0001	0.0037	0.1193
INTER14	0.10585	0.05537	0.03218
	0.0001	0.0033	0.0876
INTER15	0.06099	0.03034	0.00764
	0.0012	0.1073	0.6852
INTER16	0.14975	0.05095	0.03845
	0.0001	0.0068	0.0412
INTER17	-0.00842	0.00025	0.00180
	0.6550	0.9896	0.9240
INTER18	-0.05198	-0.01436	-0.02373
	0.0058	0.4460	0.2077

INTER19	0.00629 0.7384	-0.02271 0.2280	0.01637 0.3849
INTER20	-0.05623 0.0028	-0.02818 0.1347	-0.02189 0.2452
INTER21	-0.08291 0.0001	-0.03788 0.0443	-0.02863 0.1266

Note: The first number is the correlation between the predictor and the criterion, the second number is the significance level.

APPENDIX C
REGRESSION ANALYSIS PROGRAMS

Regression analysis attempts to predict or explain the values of the criterion variable with one or more predictor variables. The following sections expand upon the discussion of regression analysis presented in Chapter IV.

A. REQUIREMENTS AND ASSUMPTIONS

When conducting regression analysis, certain requirements must be met or assumed. One of these requirements is the use of quantitative variables.[5] Application of regression procedures also requires normality (the value of the dependent variable must be normally distributed at each value of the independent variable), homoscedasticity (the variation around the regression line must be constant for all values of the independent variable), and independence of error (the residual difference between an observed and predicted value of the dependent variable must be independent for each value of the predictor variable). Another requirement of linear regression is that a straight-line or linear relationship exist between each independent variable and the dependent variable. For purposes of this study, and based on initial investigation, these requirements are assumed to be met. However, an extensive effort to evaluate these assumptions by transforming the variables or employing complex statistical analysis packages has not been conducted.

----

[5]The inclusion of qualitative or categorical variables in regression models may be accommodated through the use of dummy variables.

B. STEPWISE REGRESSION

The SAS Stepwise process considers each of the candidate independent variables for inclusion in the model by determining the contribution the variable makes to the model. This determination is accomplished by calculating the partial F statistic for the variable, and adding it to the model if it meets the specified entry significance level. After a variable is added, the stepwise method then looks at all the variables in the model and deletes any variable that does not provide an F statistic sufficient to meet the specified significance level for remaining in the model. This process of adding and deleting variables continues until none of the variables has an F statistic significant to enter or leave the model.[6] [Ref. 12]

C. LINEAR REGRESSION

Simple linear regression is concerned with finding the statistical model or equation that best "fits" the original data. This is accomplished by defining a straight line that minimizes the differences between the actual value of the dependent variable and the value that would be predicted from the fitted line of regression. The SAS Regression procedure uses a mathematical technique, the least-squares method, to produce such an equation for the best linear model. This equation provides the intercept and slope of the sample predictor variable. With multiple linear regression, these slopes represent the unit change in the dependent variable per unit change in the independent variable, taking into account the effects of the other independent variables, and are referred to as net regression coefficients. The sample regression coefficients of the predictor

---

[6]This study used the SAS Stepwise default significance level of .15 for variables to enter or remain in the model.

variables are then used as estimates of the respective population parameters. For illustration, the program used to validate Model A is provided in Table XVI.

TABLE XVI

Sample Validation Program

```
//ADVALID JOB (2807,C110),'D OSLUND, SMC 1763',CLASS=B
//*MAIN ORG=NPGVM1.2807P
//  EXEC SAS
//FILEIN DD  DISP=SHR,DSN=MSS.S2807.ADALL4
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

* THIS PROGRAM CALCULATES THE VALIDITY OF A REGRESSION
  MODEL THROUGH THE USE OF CROSS-VALIDATION AND DOUBLE
  CROSS-VALIDATION TECHNIQUES.;

DATA DATA1;
   SET FILEIN.ADALL4;

* THE RANDOM VARIABLE CREATED IN 'ADNEWVAR' IS NOW USED
  TO SPLIT THE DATA APPROXIMATELY IN HALF.  'DERIVA' IS
  THE DERIVATION SAMPLE AND 'VALIDA' IS THE HOLD-OUT OR
  VALIDATION SAMPLE.;

DATA DERIVA;
   SET DATA1;
   IF RANDALL1 = 1;
DATA VALIDA;
   SET DATA1;
   IF RANDALL1 = 0;

* A BLOCK REGRESSION IN NOW RUN ON DERIVA TO COMPUTE AND
  OUTPUT THE PARAMETER ESTIMATES (BETAS) THAT RESULT
  FROM THE REGRESSION.  THE BETAS ARE WRITTEN TO THE DATA-
  SET WORK.BETAD.  THE MODEL IS GIVEN THE LABEL 'TAFMHATV';

PROC REG DATA=DERIVA OUTEST=BETAD;
   TAFMHATV:MODEL TAFMS1 = ADMINSCR TERMENLT DEPENDTS BLACK
                           HSDG     OTHER    NUSEX / STB;
TITLE REGRESSING ON DERIVA;

* THE NEXT STEP IS TO APPLY THE REGRESSION FORMULA (THE
  BETAS) TO THE DATA IN THE VALIDATION SAMPLE AND CALCULATE
  THE PREDICTED SCORE FOR EACH CASE IN VALIDA.  THE PRED-
  ICTED SCORES ARE WRITTEN TO WORK.PREDTAFV.  SAS USES THE
  MODEL LABEL (TAFMHATV) AS THE VARIABLE NAME FOR THE VALIDA
  PREDICTED SCORES.  THE SCORE PROCEDURE PRODUCES NO
  PRINTED OUTPUT.;

PROC SCORE OUT=PREDTAFV TYPE=OLS SCORE=BETAD
       DATA=VALIDA PREDICT;
   VAR ADMINSCR TERMENLT DEPENDTS BLACK
       HSDG     OTHER    NUSEX;

* THE FIRST VALIDITY COEFFICIENT IS NOW CALCULATED BY FIND-
  ING THE CORRELATION BETWEEN VALIDA'S ACTUAL SCORES AND
  VALIDA'S PREDICTED SCORES.;

PROC CORR DATA=PREDTAFV;
     VAR TAFMS1 TAFMHATV;
TITLE FIRST VALIDITY COEFFICIENT.;
```

```
* NOW TO REPEAT THE PROCESS TO UTILIZE THE DOUBLE CROSS-
  VALIDATION TECHNIQUE.  THIS TIME A REGRESSION IS RUN
  ON VALIDA AND THE RESULTING BETAS (BETAV) ARE USED TO
  PREDICT THE SCORES OF THE CASES IN DERIVA.  DERIVA'S
  ACTUAL AND PREDICTED SCORES ARE THEN CORRELATED TO
  FIND THE SECOND VALIDITY COEFFICIENT.;

PROC REG DATA=VALIDA OUTEST=BETAV;
   TAFMHATD:MODEL TAFMS1 = ADMINSCR TERMENLT DEPENDTS BLACK
                           HSDG     OTHER    NUSEX / STB;
TITLE REGRESSING ON VALIDA;

PROC SCORE OUT=PREDTAFD TYPE=CIS SCORE=BETAV
        DATA=DERIVA PREDICT;
   VAR ADMINSCR TERMENLT DEPENDTS BLACK
       HSDG     OTHER    NUSEX;

PROC CORR DATA=PREDTAFD;
     VAR TAFMS1 TAFMHATD;
TITLE SECOND VALIDITY COEFFICIENT;

/*
//
```

APPENDIX D

DISCRIMINANT ANALYSIS PROGRAMS

Discriminant Analysis allows observations to be classified into two or more groups on the basis of one or more numeric variables. The following sections expand upon the discussion of discriminant analysis presented in Chapter IV. For illustration, Table XVII shows the program used to produce the classification matrices for the derivation and validation samples for Model A.

A. REQUIREMENTS AND ASSUMPTIONS

As was the case with regression analysis, discriminant analysis also requires that certain basic assumptions be met. First, the observations in the data set should be members of two or more mutually exclusive groups. Therefore, the groups must be defined so that each case will belong to only one group. Another statistical property required of discriminating variables is that they may not be linear combinations of other variables. Thus, the sum or average of several variables may not be used along with those variables. There are three other assumptions to be considered. The population covariance matrices must be equal for each group, each group is to be drawn from a population which has a multivariate normal distribution, and discriminating variables must be measured at the interval or ratio levels. Ideally, these variables will be continuous, but they need not be. [Ref. 17] This study assumes these requirements have been met. However, an effort to evaluate these properties was not conducted since, in practice, the discriminant analysis technique is rather robust and can tolerate some deviation from these assumptions [Ref. 18].

E. DISCRIMINANT ANALYSIS

The first step of discriminant analysis is to weight and linearly combine the discriminating variables so that the groups will be as statistically distinct as possible. The derived equations, called discriminant functions, combine the group characteristics using a measure of generalized squared distance[7] that will allow one to identify the group to which a case belongs or most closely resembles.

The classification process may assume that membership in a group has equal likelihood of occurring. However, it may be more desirable to incorporate the prior probability of group membership into the classification function to improve prediction accuracy or minimize the cost of prediction errors. In this study, membership in a success group was on the order of 80 percent. Therefore, it was appropriate to consider prior probabilities so that those cases predicted as unsuccessful would be classified as such only if strong evidence exists that they belong there.

The ultimate concern in developing a mathematical model is that it predict well or provide a reasonable description of the real world. Once a model is developed which provides satisfactory discrimination for cases of group membership, classification functions may be derived and applied to the classification of new cases with unknown group membership. A good test of the adequacy and accuracy of the discriminant model is the percentage of correct classifications, commonly called the "hit-rate". This test is accomplished by applying the classification function to the known cases from which the function was derived. The percentage of correctly

---

[7]The procedure conducted a likelihood ratio test of homogeneity of the within-group covariance matrices for each model. This test was statistically significant for each model. Therefore, the within-group matrices were used as the basis of the measure of generalized squared distance in developing the classification criterion. [Ref. 12]

classified cases provides an indication of the accuracy of the procedure and indirectly confirms the degree of group separation. The results may be depicted in a classification matrix.

When the sample size is large enough, as it is in this study, a further check of the classification accuracy may be conducted by randomly splitting the sample into two subsets. The classification function is derived on one subset and validated on the other subset. A comparison of the two hit-rates provides the measure of accuracy of the model. [Ref. 17]

TABLE XVII

Sample Discriminant Analysis Program

```
//DISCPGMS JOB (2807,C110),'D OSLUND, SMC 1763',CLASS=B
//*MAIN ORG=NPGVM1.2807P
//  EXEC SAS
//FILEIN DD  DISP=SHR,DSN=MSS.S2807.ADALL4
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

* THIS PURPOSE OF THIS PROGRAM IS TO ALLOW THE VALIDITY
  OF A DISCRIMINANT MODEL TO BE INVESTIGATED.  A CLASS-
  IFICATION FUNCTION IS DERIVED FROM THE DERIVA SAMPLE
  AND THIS FUNCTION IS USED TO CLASSIFY THE CASES IN THE
  VALIDATION (OR HOLD-OUT) SAMPLE.  THE TWO CLASSIFICATION
  MATRICIES ARE THEN USED TO ALLOW THE 'HIT RATE' ON EACH
  SAMPLE TO BE CALCULATED.;

DATA DATA1;
   SET FILEIN.ADALL4;

* USING THE RANDOM VARIABLE TO SPLIT THE SAMPLE APPROXIM-
  ATELY IN HALF.;

DATA DERIVA;
   SET DATA1;
   IF RANDALL1=1;
DATA VALIDA;
   SET DATA1;
   IF RANDALL1=0;

* CALCULATING THE CLASSIFICATION MATRIX FOR DERIVA AND
  WRITING OUT THE CLASSIFICATION FUNCTION DERIVED FROM
  DERIVA TO WORK.D.;

PROC DISCRIM DATA=DERIVA OUT=D POOL=TEST;
   CLASS SUCCTAF;
   VAR DEPENDTS HSDG     BLACK    TERMENLT
       NUSEX    OTHER    ADMINSCR;
   PRIORS PROPORTIONAL;
TITLE DISCRIM ON DERIVA.;

* NOW THE CLASSIFICATION FUNCTION FROM DERIVA IS USED TO
  CLASSIFY THE CASES IN VALIDA.;

PROC DISCRIM DATA=D TESTDATA=VALIDA;
   TESTCLASS SUCCTAF;
TITLE DERIVA''S FUNCTION APPLIED TO VALIDA.;

/*
//
```

APPENDIX E

UTILITY ANALYSIS PROGRAMS

This appendix provides further details of the information contained in Chapter V, and gives examples of the SAS programs and outputs.

A. CALCULATION OF CELL PROBABILITIES

The method used to calculate cell probabilities in this study depends on whether a regression or a discriminant model is being evaluated. A regression model can be viewed simply as a formula for calculating predicted scores, whereas a discriminant model actually classifies cases as predicted successes or predicted failures. Because of this difference, the calculation of cell probabilities is more complicated for regression models than for discriminant models.

1. Regression Models

A regression model and the data from which it was developed provide information on the predicted and actual scores for each case. In order to classify these cases into the four selection outcomes, the cut score on the predictor and the score on the criterion above which people are considered to be successful must be known. If the criterion is constructed as a dichotomous (success/fail) variable, then the cases assigned a value of "one" are considered successful and those with a value of "zero" are considered unsuccessful. If the criterion is a continuous variable (such as length of service) then a value on the scale must be chosen as the dividing line between success and failure.

The choice of the cut score is not such a simple matter, and cannot be arbitrarily assigned as can the distinction between success and fail. The choice of the cut score, as mentioned before, often depends on the desired selection ratio. In the absence of information on the desired selection ratio, cell probabilities are calculated for each of many possible cut scores, and a cut score is eventually chosen based on which set of cell probabilities maximizes the utility of the model. In a data set containing actual and predicted scores, different sets of cell probabilities can be calculated if each predicted score is considered to be a cut score. Table XVIII contains five pairs of actual and predicted scores which will be used to illustrate the method.

TABLE XVIII

Illustrative Actual and Predicted Scores

Actual Criterion Score	Predicted Criterion Score
50	44
44	46
49	47
46	49
49	50

In this illustration, cases who serve 48 months or longer are considered to be successful. Each different predicted score will be considered as a cut score and cell counts for each cut score will be calculated. If the cut score is 44 months, then all cases with a predicted score of 44 months or more will be accepted, and those with a predicted score of less than 44 months will be rejected. In this example, for a cut score of 44, all cases will be accepted. No one is rejected, therefore, valid negatives and false negatives will be zero. Of the five cases

AD-A152 127 THE DEVELOPMENT OF AN ENLISTMENT STANDARDS MODEL FOR THE NAVY AVIATION MACHINIST'S MATE (AD) RATING(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D A OSLUND ET AL. JUN 84

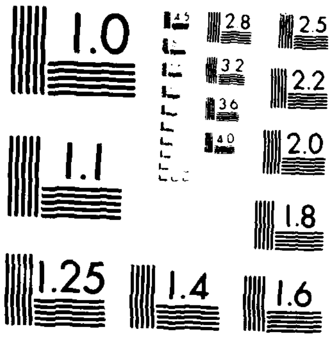
2/2

UNCLASSIFIED F/G 5/9 NL

END

FILMED

DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

accepted, three have actual LOS of 48 months or more (successes). Therefore, the number of valid positives is three. Two of the five cases accepted had actual LOS of less than 48 months (failures). Therefore, false positives will be two. Thus the first set of cell probabilities that result when the cut score is 44 are: PVP = 3/5, PFP = 2/5, PFN = 0 and PVN = 0. The next set of cell probabilities will result when 46 months is considered to be the cut score. One case had a predicted LOS of less than 46, Therefore, he would be rejected. His actual LOS is 50 months, so he was falsely rejected, i.e. FN = 1. No one else was rejected so VN = 0. Four cases had a predicted LOS of 46 or greater so all four would be accepted. Of these four, two had actual LOS of less than 48 months (FP), and two had actual LOS of 48 months or more (VP). Thus for a cut score of 46, PVP = 2/5, PFP = 2/5, PFN = 1/5 and PVN = 0. This process is repeated until five sets of cell probabilities (one for each different predicted score) are calculated.

2. Discriminant Models

In a discriminant model the criterion is a categorical (0,1) variable. The output from the SAS Discriminant procedure is a two by two table where the cases are predicted to be either a "zero" or a "one", and the prediction is compared to the actual score. Table XIX gives an abbreviated example of the output from the discriminant procedure.

The columns are the model's predicted scores for the 750 cases in this hypothetical sample. Here the model predicts that 300 of the cases will score "zero" on the criterion, and it predicts that 450 of the cases will score a "one" on the criterion. The rows are the actual scores of the cases. 250 people actually scored "zero" (failures) and

TABLE XIX

Illustrative Discriminant Example

		Predicted		
		0	1	Total
Actual	0	100	150	250
	1	200	300	500
	Total	300	450	750

500 people actually scored "one" (successes). Because, in effect, the discriminant procedure chooses its own cut score, the four cell probabilities can be derived directly from the output. The predicted "ones" are people that the model classifies as accept. Of these 450, 150 actually failed so they are false positives, and the remaining 300 were successful, so they are valid positives. Of the 300 cases that the model would have rejected (predicted "zeros"), 100 were failures (valid negatives) and 200 were successes (false negatives). Again the cell probabilities are found by dividing each count by the number of cases. Therefore, PVP = 300/750, PFP = 150/750, PFN = 200/750 and PVN = 100/750. For a discriminant model, there is only one set of cell probablities to be calculated.

E. ESTIMATION OF CELL UTILITIES

In order to calculate the overall utility of a model, utilities associated with each selection outcome need to be estimated. "Although the assignment of utility values to outcomes may very well be the 'Achilles Heel' of decision theory, it is not a problem that can be ignored by any institution that makes personnel decisions." [Ref. 19]

Ideally each selection outcome should have a uniquely estimated utility. Because of the difficulty in estimating

utilities for each outcome (particularly for the false and valid negatives), relative utilities are estimated. It is apparent that a person who is correctly selected (valid positive) has a positive worth to the organization. A reasonable estimate of this worth is the marginal product of the employee. In this study it is assumed that the navy compensates sailors at the full value of their marginal product, and the Billet Cost Model provides an estimate of the cost to the Navy of staffing a billet [Ref. 16]. Because relative utilities are the issue at this time, the utility of a valid positive (U1) is assigned the value of +1.

It is a reasonable assumption that the utility of a false positive is a negative number. As the employee was not judged to be successful, his marginal product was probably less than the marginal cost to keep him in the job. In addition a poor performer may adversely affect the performance and productivity of his fellow employees, and when he leaves, additional expense is necessary to find a replacement. On the other hand, it is unlikely that a poor performer does not contribute anything to the organization, and thus it is obviously difficult to estimate the magnitude of the disutility of a false positive. In this study a minor form of sensitivity analysis is undertaken to circumvent this estimation difficulty, and expected overall utilities are calculated for three different relative values of false positive utility (U2). These values are -.5, -1, and a relatively extreme assumption, -2.

The disutility of a false negative is also difficult to estimate, partly because it is not known what happens to the applicant after he is rejected. If the Navy rejects an applicant to the AD rating but accepts him in another rating where he is subsequently successful, then his disutility could be reasonably argued to be zero. If, however, the

Navy rejects him altogether when he would have been successful if selected, then the costs of attracting and testing him are wasted and additional costs are required to attract and test another applicant. These costs will depend on the state of the recruiting market at the time. If there are many good quality applicants then the disutility of rejecting a potentially successful applicant may be small. Again, as a type of sensitivity analysis, three relative values for the utility of a false negative (U3) are considered: 0, -.25 and -.5.

It is not obvious that any utility should be assigned to U4, the utility of a valid negative. The person would have failed anyway, so nothing was gained by rejecting him. However, when viewed from an economist's viewpoint in relation to opportunity costs, the fact that the person was correctly rejected means that the organization did not have to bear the cost of incorrectly accepting someone who turns out to be unsuccessful. Thus, correctly rejecting an applicant is of equal and opposite utility to incorrectly accepting him. Therefore, U4 = -U2.

The use of relative utilities is a convention to simplify the estimation of cell utilities. In the above discussion relative utilities are estimated on the basis that the utility of a valid positive is +1. However, the values of U1 through U4 that are used in the formula for overall expected utility, (Equation 5.1), need to be expressed in actual dollars. As mentioned above, the Billet Cost Model is used to estimate the utility of a valid positive. The standard manyear cost of an E-4 Aviation Machinist's Mate is $24,163. This cost comes from financial year 1983 data and represents the total cost to the Navy of creating and filling a job slot over one full year. [Ref. 16] A utility of +1 is therefore equivalent to +$24,163, a utility of -.5 will be -$12,082, and so on.

C. PROGRAMS USED TO CALCULATE UTILITIES

As mentioned in Section A. above, the calculation of cell probabilities for a regression model is a fairly tedious and repetitive procedure. This section contains three sample programs used to calculate the expected utility of a model. Explanatory comments are provided following each set of SAS statements. The first program (Table XX) computes the predicted criterion score for each case and writes the results out to a file called "RTYHATA". Table XXI shows part of the output from the first program. The second program's main purpose (Table XXII) is to calculate the cell probabilities that would result if each different predicted score were used as a cut score. The cell probabilities are written out to a file called "RTUTILA". The program also calculates the expected utilities for one set of cell utilities and outputs the 30 largest utilities that result (Table XXIII). The third program (Table XXIV) calculates the utilities for six different sets of cell utilities.

As explained before, only one set of probabilities results from a discriminant model and these can be readily gained from the discriminant output. No programs were used to calculate the expected utilities of a discriminant model and these calculations were done by hand.

D. CALCULATION OF BASE LINE UTILITIES

As described in Chapter V, the utility of the Navy's original selection strategy (the base line utility) needs to be calculated in order for comparisons to be made. Observation 4 in Table XXIII demonstrates that when all the cases are accepted (41.0774 is the lowest predicted score), the selection ratio is obviously 1 and PVP = .860638 (which is the base rate) and PFP = 1 - PVP = .139362. No one is

rejected, therefore PFN and FVN are zero. The expected utility under these circumstances is:

$$EU = .860638(\$24,163) + .139362(-\$12,082) + 0 + 0 = \$19,112$$

As Table XXIII shows, the maximum utility occurs when the cut score is slightly higher than the lowest predicted score (there are five cases with a predicted score of less than 43.2692 in Table XXI). This maximum utility ($19,135) is .12 percent greater than the base line utility of $19,112.

TABLE XX

First Utility Analysis Program

```
//SEICTIII1 JOB (2840,C104),'SEI CLARK, SMC 1560',CLASS=B
//*MAIN ORG=NPGVM1.2840P
//   EXEC SAS
//FILEIN DD  DISP=SHR,DSN=MSS.S2807.ADALL4
//FILEOUT DD UNIT=3330V,MSVGP=PUB4A,DISP=(NEW,CATLG,DELETE),
//         DSN=MSS.S2840.RTYHATA,
//            DCB=(BLKSIZE=6400)
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

*   THE PURPOSE OF THIS PROGRAM IS TO CALCULATE THE PREDICTED
    SCORE FOR EACH CASE (USING THE MODEL DEVELOPED PRE-
    VIOUSLY), AND TO WRITE OUT THE ACTUAL AND PREDICTED
    SCORES TO A FILE IN MASS STORAGE.;

DATA DATA1;
   SET FILEIN.ADALL4;
   RENAME TAFMS1=Y;

*    RENAMING THE CRITERION VARIABLE;

PROC REG DATA=DATA1 OUTEST=BETAS;
  YHAT:MODEL Y =
  DEPENDTS HSDG BLACK OTHER NUSEX TERMENLT ADMINSCR / STB;
TITLE BLOCK REGRESSION TO OUTPUT BETAS.;

PROC SCORE OUT=PREDY TYPE=OLS SCORE=BETAS DATA=DATA1 PREDICT;
  VAR DEPENDTS HSDG BLACK OTHER NUSEX TERMENLT ADMINSCR;

*    CALCULATES THE PREDICTED SCORES, AND WRITES THEM TO
     DATASET 'PREDY'.
     NOTE:  THE SCORE PROCEDURE TAKES THE MODEL LABEL (YHAT)
            AND USES THAT LABEL AS THE VARIABLE NAME FOR THE
            PREDICTED SCORE. ;

DATA PREDY2;
   SET PREDY;
   KEEP YHAT Y SUCCTAF;

PROC SORT DATA=PREDY2 OUT=FILEOUT.RTYHATA;
   BY YHAT;

*    SORTS THE OUTPUT FILE INTO ASCENDING YHAT ORDER,
     AND WRITES OUT THE SORTED DATA TO MASS STORAGE.;

DATA TEST;
  SET FILEOUT.RTYHATA;
  IF _N_ LE 10 OR (_N_ GT 1270 AND _N_ LE 1280)
               OR _N_ GT 2790;

PROC PRINT DATA=TEST SPLIT=*;
  LABEL     Y=ACTUAL*CRITERION*SCORE
         YHAT=PREDICTED*CRITERION*SCORE
      SUCCTAF=SUCCESS ON*CRITERION;
TITLE THE FIRST 10, MIDDLE 10 AND LAST 30 OBS OF RTYHATA;
TITLE2;
TITLE3 NOTE:  SORTED IN ASCENDING ORDER OF YHAT.;

PROC UNIVARIATE DATA=FILEOUT.RTYHATA PLOT;
  VAR YHAT Y SUCCTAF;
TITLE STATS OF THE ACTUAL AND PREDICTED CRITERION SCORES;
```

TABLE XXI

Partial Output from the First Utility Program

THE FIRST 10, MIDDLE 10 AND LAST 30 OBS OF RIYHATA

NOTE: SORTED IN ASCENDING ORDER OF YHAT.

OBS	ACTUAL CRITERION SCORE	SUCCESS ON CRITERION	PREDICTED CRITERION SCORE
1	45	0	41.0774
2	23	0	41.0774
3	66	1	42.0297
4	12	0	42.0297
5	16	0	42.8960
6	21	0	43.2692
7	11	0	43.2692
8	54	1	43.2692
9	48	1	43.2692
10	48	1	43.2692
11	50	1	48.7601
12	51	1	48.7601
13	47	1	48.7601
14	48	1	48.7601
15	65	1	48.7601
16	71	1	48.7601
17	48	1	48.7601
18	33	0	48.7601
19	48	1	48.7601
20	49	1	48.7601
21	48	1	56.8469
22	51	1	56.8469
23	58	1	56.8469
24	59	1	56.8469
25	52	1	56.8469
26	70	1	56.8469
27	51	1	56.8469
28	60	1	56.8469
29	49	1	56.8469
30	50	1	56.8469
31	56	1	57.7992
32	60	1	57.7992
33	54	1	57.7992
34	27	0	57.7992
35	36	0	57.7992
36	54	1	57.7992
37	52	1	57.7992
38	63	1	57.7992
39	68	1	57.7992
40	54	1	57.7992
41	51	1	57.7992
42	66	1	57.7992
43	64	1	57.7992
44	64	1	57.7992
45	72	1	57.7992
46	63	1	58.0774
47	64	1	59.1512
48	52	1	59.3928
49	62	1	61.2115
50	46	1	61.2115

TABLE XXII

Second Utility Analysis Program

```
//SEIDTII2 JOB (2840,C104),'SEI CLARK, SMC 1560',CLASS=E
//*MAIN ORG=NPGVM1.2840P
// EXEC SAS
//SAS.WORK DD SPACE=(CYL,(12,4))
//FILEIN DD  DISP=SHR,DSN=MSS.S2840.RTYHATA
//FILEOUT DD UNIT=3330V,MSVGP=PUB4A,DISP=(NEW,CATLG,DELETE),
//         DSN=MSS.S2840.RTUTILA,
//            DCB=(BLKSIZE=6400)
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

* THE PURPOSE OF THIS PROGRAM IS TO WRITE OUT A FILE TO MASS
  STORAGE WHICH CONTAINS THE VALUES OF PVP, PFP, PFN AND PVN
  THAT RESULT WHEN EACH PREDICTED SCORE IS USED TO SEPARATE
  THE CASES INTO ACCEPT AND REJECT GROUPS (IE. OUTPUT THE
  CELL PROBABILITIES THAT RESULT WHEN EACH PREDICTED SCORE
  IS USED AS A CUTTING SCORE).

  THE INPUT FILE CONTAINS 3 VARIABLES, AND THE OBSERVATIONS
  (OR CASES) ARE SORTED IN ASCENDING ORDER OF 'YHAT'.  YHAT
  IS THE PREDICTED LOS (FROM THE MODEL DEVELOPED EARLIER) OF
  EACH CASE, 'Y' IS THE ACTUAL LOS IN MONTHS AND 'SUCCTAF'
  IS A DUMMY VARIABLE WHERE EACH CASE IS CATEGORIZED AS A
  SUCCESS (1) OR AS A FAILURE (0).;

DATA DATA1;
   SET FILEIN.RTYHATA;
   DROP Y;
   RENAME SUCCTAF = Y;

* THE DATA IS READ IN AND THE ACTUAL LOS IN MONTHS VARIABLE
  IS DROPPED AND THE DUMMY VARIABLE IS RENAMED 'Y'.;

PROC SUMMARY DATA=DATA1;
   VAR Y;
   OUTPUT OUT=DATA2 SUM=NSUCC N=NCASE;

* HERE THE NUMBER OF SUCCESSFUL CASES IN THE DATA (NSUCC) IS
  FOUND BY SUMMING THE 1'S AND 0'S IN VARIABLE 'Y'.  ANOTHER
  VARIABLE 'NCASE' IS CREATED WHICH IS THE NUMBER OF CASES
  IN THE DATA.  THESE TWO VARIABLES (EACH A SINGLE NUMBER)
  ARE WRITTEN TO DATA SET WORK.DATA2.;

DATA DATA3;
   IF _N_ EQ 1 THEN SET DATA2;
   NFAIL = NCASE-NSUCC;
   SET DATA1;

* THE VARIABLES NCASE, NSUCC AND NFAIL (THE NUMBER OF UNSUC-
  CESSFUL CASES IN THE DATA) ARE ADDED TO DATA1.  NCASE,
  NSUCC AND NFAIL ARE EACH SINGLE NUMBERS THAT ARE REPEATED
  FOR EACH OBSERVATION.  EG. NCASE IS A COLUMN OF 500'S
  (SAY), NSUCC IS A COLUMN OF 325'S (SAY) AND THEREFORE
  NFAIL IS A COLUMN OF 175'S.;
```

```
DATA DATA4;
   SET DATA3;
   U1= 24163; U2= -12082; U3= -6041; U4= 12082;
   RETAIN NZERO 0;
   RETAIN LASTYHAT 0;
   IF YHAT NE LASTYHAT THEN LINK CALCS; ELSE LINK ZEROS;
   IF Y=0 THEN NZERO=NZERO+1;
   LASTYHAT=YHAT;
   RETURN;
   CALCS: VP = NSUCC-(_N_-1-NZERO);
          FP = NFAIL-NZERO;
          FN = _N_-1-NZERO;
          VN = NZERO;
          UTIL = (U1*VP + U2*FP + U3*FN + U4*VN)/NCASE;
          SRATIO = (VP+FP)/NCASE;
          SUCCRATE = VP/(VP+FP);
          RETURN;
   ZEROS: VP = 0;      FP = 0;        FN = 0;          VN = 0;
          UTIL = 0;  SRATIO = 0;  SUCCRATE = 0;
          RETURN;

* THIS IS THE HEART OF THE PROGRAM WHERE SUBTLE LOGIC IS
  EMPLOYED. 'NZERO' IS A COUNTER WHICH COUNTS THE NUMBER OF
  0'S IN THE 'Y' VARIABLE DOWN TO AND INCLUDING THE LINE (OR
  OBSERVATION) CONTAINING THE 'CURRENT' CUTTING SCORE.
  FOR EXAMPLE, IF THERE ARE 150 ZEROS AND 250 ONES AMONG THE
  FIRST 400 OBS. OF 'Y', THEN THE 400TH OBS. OF 'NZERO' WILL
  BE 150.  IF THE 401ST OBS. OF 'Y' IS A ZERO THEN THE 401ST
  OBS OF 'NZERO' WILL BE 151.  TO CONTINUE THE EXAMPLE,
  BECAUSE THE INPUT DATA IS SORTED IN ASCENDING ORDER OF
  'YHAT', THE 400 CASES PRECEDING THE 401ST CASE (WHICH IS
  THE CURRENT CUTTING SCORE), WOULD ALL BE CLASSIFIED AS
  REJECT BECAUSE THEIR PREDICTED SCORE IS LESS THAN THE
  CUTTING SCORE.  THE 400TH OBS. OF 'NZERO' TELLS US HOW
  MANY OF THESE REJECTED CASES WERE FAILURES AND THEREFORE
  VN = NZERO.  'NFAIL' IS THE TOTAL NUMBER OF CASES THAT
  FAILED, THEREFORE NFAIL-VN (SAME AS NFAIL-NZERO) = FP.
  THE NUMBER OF ONES IN THE REJECTED 400 CASES (FN) IS THE
  CURRENT OBS. (401), MINUS 1, MINUS THE NUMBER OF ZEROS, OR
  FN = 401-1-150 = 250.  FINALLY, 'NSUCC' IS THE TOTAL
  NUMBER OF SUCCESSES, THEREFORE NSUCC-FN IS THE VALUE OF VP.

  'LASTYHAT' IS USED TO PRECLUDE ANY ERRORS THAT WOULD BE
  GENERATED WHEN TWO OR MORE VALUES OF 'YHAT' ARE IDENTICAL.
  IF THE NEXT POTENTIAL CUTTING SCORE IS THE SAME AS THE
  PREVIOUS ONE, THEN NO CELL PROBABILITIES, ETC ARE CALCUL-
  ATED, AND ZEROS ARE ASSIGNED.
  NOTE:  DUE TO THE USE OF THE KEYWORD 'RETAIN', THE VALUES
         OF NZERO AND LASTYHAT USED IN THE CALCULATIONS AND
         IN THE FIRST 'IF' STATEMENT ARE THE VALUES FROM THE
         PREVIOUS OBSERVATION.

  THE DATA STEP ALSO INITIALIZES A SET OF INDIVIDUAL CELL
  UTILITIES (U1 - U4) AND CALCULATES THE OVERALL UTILITY
  ASSOCIATED WITH EACH CUTTING SCORE.  ALSO THE SELECTION
  RATIO AND THE SUCCESS RATE RESULTING FROM EACH CUTTING
  SCORE ARE CALCULATED.;

DATA DATA5;
   SET DATA4;
   PVP = VP/NCASE;     PFP = FP/NCASE;
   PFN = FN/NCASE;     PVN = VN/NCASE;
   KEEP YHAT UTIL PVP PFP PFN PVN SRATIO SUCCRATE;
   RENAME YHAT = CSCORE;
LABEL
CSCORE=CUT SCORE ON PREDICTOR;

* CONVERTING THE CELL COUNTS TO PROBABILITIES.;
```

```
PROC SORT DATA=DATA5 OUT=FILEOUT.RTUTILA;
   BY DESCENDING UTIL;

* SORTING BY UTIL BEFORE WRITING OUT THE PREVIOUSLY KEPT
  VARIABLES TO A FILE IN MASS STORAGE.;

DATA DATA6;
   SET FILEOUT.RTUTILA;
   IF _N_ LE 30;

PROC PRINT DATA=DATA6;
TITLE THE 30 LARGEST UTILITIES IN THE FILEOUT.;
TITLE2;
TITLE3 THE BASE UTILITY IS 19112, AND THE;
TITLE4;
TITLE5 BASE LINE SUCCESS RATE IS 0.8606;

PROC PLOT DATA=DATA6;
   PLOT UTIL * CSCORE = '+' / VREF =19112;
TITLE THE TOP 30 UTILITIES PLOTTED AGAINST CUTTING SCORE.;
TITLE2;
TITLE3 NOTE:  THE HORIZ. LINE IS THE BASE LINE UTILITY,;
TITLE4;
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY''S;
TITLE6;
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);

PROC PLOT DATA=DATA6;
   PLOT UTIL * SRATIO = '+' / VREF =19112;
TITLE THE TOP 30 UTILITIES PLOTTED AGAINST SELECTION RATIO.;
TITLE2;
TITLE3 NOTE:  THE HORIZ. LINE IS THE BASE LINE UTILITY,;
TITLE4;
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY''S;
TITLE6;
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);

PROC PLOT DATA=FILEOUT.RTUTILA;
   PLOT UTIL * SRATIO = '+' / VREF =19112;
TITLE PLOTTING ALL UTILITIES AGAINST SELECTION RATIO.;
TITLE2;
TITLE3 NOTE:  THE HORIZ. LINE IS THE BASE LINE UTILITY,;
TITLE4;
TITLE5 IE. THE UTILITY RESULTING FROM THE NAVY''S;
TITLE6;
TITLE7 ORIGINAL SELECTION STRATEGY. (19112);

/*
//
```

TABLE XXIII

Partial Output from the Second Utility Program

THE 30 LARGEST UTILITIES IN THE FILEOUT.

THE BASE UTILITY IS 19112, AND THE

BASE LINE SUCCESS RATE IS 0.8606.

OBS	CSCORE	UTIL	SRATIO	SUCCRATE	PVP	PFP	PFN	PVN
1	43.2692	19135	0.998	0.862	0.860	0.138	0.001	0.001
2	42.0297	19129	0.999	0.861	0.861	0.139	0.001	0.001
3	42.8960	19127	0.998	0.862	0.860	0.138	0.001	0.001
4	41.0774	19112	1.000	0.861	0.861	0.139	0.000	0.000
5	43.8483	18998	0.988	0.864	0.853	0.135	0.007	0.005
6	44.2215	18988	0.988	0.863	0.853	0.135	0.008	0.005
7	44.6648	18865	0.982	0.863	0.848	0.134	0.012	0.005
8	45.0879	18855	0.982	0.863	0.848	0.134	0.013	0.005
9	45.6261	18791	0.980	0.863	0.846	0.134	0.015	0.005
10	46.0401	18780	0.979	0.863	0.845	0.134	0.015	0.005
11	46.5683	18748	0.978	0.863	0.844	0.134	0.016	0.005
12	46.6815	16614	0.856	0.877	0.751	0.105	0.110	0.034
13	47.2790	16591	0.855	0.877	0.750	0.105	0.111	0.034
14	47.5206	16514	0.851	0.877	0.747	0.105	0.114	0.035
15	48.2312	15725	0.807	0.883	0.713	0.095	0.148	0.045
16	48.3870	15719	0.805	0.884	0.712	0.094	0.149	0.046
17	48.7601	15535	0.799	0.883	0.705	0.093	0.155	0.046
18	49.3392	6169	0.395	0.896	0.353	0.041	0.507	0.098
19	49.4708	5961	0.383	0.899	0.345	0.039	0.516	0.101
20	49.5373	5300	0.359	0.896	0.322	0.037	0.539	0.102
21	49.7124	5289	0.358	0.896	0.321	0.037	0.539	0.102
22	49.9806	1834	0.217	0.898	0.195	0.022	0.666	0.117
23	50.4231	1756	0.213	0.899	0.192	0.022	0.669	0.118
24	50.5788	1298	0.193	0.904	0.174	0.018	0.686	0.121
25	50.9329	195	0.146	0.910	0.133	0.013	0.727	0.126
26	51.5310	182	0.145	0.912	0.133	0.013	0.728	0.127
27	41.0774	0	0.000	0.000	0.000	0.000	0.000	0.000
28	42.0297	0	0.000	0.000	0.000	0.000	0.000	0.000
29	43.2692	0	0.000	0.000	0.000	0.000	0.000	0.000
30	43.2692	0	0.000	0.000	0.000	0.000	0.000	0.000

TABLE XXIV

Third Utility Analysis Program

```
//SEIUTIL3 JOB (2840,C104),'SEI CLARK, SMC 1560',CLASS=E
//*MAIN ORG=NPGVM1.2840P
//   EXEC SAS
//FILEIN DD  DISP=SHR,DSN=MSS.S2840.RTUTILA
//SYSIN DD *
OPTIONS LS=80 NOCENTER;

*  THIS PROGRAM EXPLORES THE EFFECTS OF USING DIFFERENT CELL
   UTILITIES FOR THE CALCULATION OF OVERALL UTILITY.;

DATA DATA1;
   SET FILEIN.RTUTILA;

   U1A= 1   ; U2A= -.5 ; U3A=  0  ; U4A=  .5  ;
   U1B= 1   ; U2B= -1  ; U3B=  0  ; U4B=  1   ;
   U1C= 1   ; U2C= -2  ; U3C=  0  ; U4C=  2   ;
   U1D= 1   ; U2D= -.5 ; U3D= -.5 ; U4D=  .5  ;
   U1E= 1   ; U2E= -1  ; U3E= -.5 ; U4E=  1   ;
   U1F= 1   ; U2F= -2  ; U3F= -.5 ; U4F=  2   ;

   UTILA= (PVP*U1A + PFP*U2A + PFN*U3A + PVN*U4A)*24163;
   UTILB= (PVP*U1B + PFP*U2B + PFN*U3B + PVN*U4B)*24163;
   UTILC= (PVP*U1C + PFP*U2C + PFN*U3C + PVN*U4C)*24163;
   UTILD= (PVP*U1D + PFP*U2D + PFN*U3D + PVN*U4D)*24163;
   UTILE= (PVP*U1E + PFP*U2E + PFN*U3E + PVN*U4E)*24163;
   UTILF= (PVP*U1F + PFP*U2F + PFN*U3F + PVN*U4F)*24163;

PROC SORT DATA=DATA1 OUT=FIRST;
   BY DESCENDING UTILA;
DATA FIRST;
   SET FIRST;
   KEEP OSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILA;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 19112 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1   , U2= -.5  , U3= 0    , U4= .5   .;
PROC PLOT DATA=FIRST;
   PLOT UTILA * SRATIO = '+' / VREF =19112;

PROC SORT DATA=DATA1 OUT=SECOND;
   BY DESCENDING UTILB;
DATA SECOND;
   SET SECOND;
   KEEP OSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILB;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 17428 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1   , U2= -1   , U3= 0    , U4= 1    .;
PROC PLOT DATA=SECOND;
   PLOT UTILB * SRATIO = '+' / VREF =17428;
```

```
PROC SORT DATA=DATA1 OUT=THIRD;
   BY DESCENDING UTILC;
DATA THIRD;
   SET THIRD;
   KEEP CSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILC;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 14061 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1     , U2= -2     , U3= 0     , U4= 2     .;
PROC PLOT DATA=THIRD;
   PLOT UTILC * SRATIO = '+' / VREF =14061;

PROC SORT DATA=DATA1 OUT=FOURTH;
   BY DESCENDING UTILD;
DATA FOURTH;
   SET FOURTH;
   KEEP CSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILD;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 19112 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1     , U2= -.5    , U3= -.5   , U4= .5    .;
PROC PLOT DATA=FOURTH;
   PLOT UTILD * SRATIO = '+' / VREF =19112;

PROC SORT DATA=DATA1 OUT=FIFTH;
   BY DESCENDING UTILE;
DATA FIFTH;
   SET FIFTH;
   KEEP CSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILE;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 17428 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1     , U2= -1     , U3= -.5   , U4= 1     .;
PROC PLOT DATA=FIFTH;
   PLOT UTILE * SRATIO = '+' / VREF =17428;

PROC SORT DATA=DATA1 OUT=SIXTH;
   BY DESCENDING UTILF;
DATA SIXTH;
   SET SIXTH;
   KEEP CSCORE PVP PFP PFN PVN SRATIO SUCCRATE UTILF;
   IF _N_ LE 30;
PROC PRINT;
TITLE BASE UTILITY IS 14061 AND BASE SUCCESS RATE IS .8606;
TITLE2;
TITLE3 U1= 1     , U2= -2     , U3= -.5 , U4= 2     .;
PROC PLOT DATA=SIXTH;
   PLOT UTILF * SRATIO = '+' / VREF =14061;

/*
//
```

LIST OF REFERENCES

1. Navy Enlisted Career Guide, 1980-81.

2. Center for Naval Analyses Report CRC 425, A New Look at Success Chances of Recruits Entering the Navy (SCREEN), by R. F. Lockman and P. M. Lurie, February 1980.

3. Center for Naval Analyses Report 81-0048, Relating Enlistment Standards to Job Performance: A Pilot Study for Two Navy Ratings, by P. M. Lurie, January 1981.

4. Center for Naval Analyses Report CRC 450, Continuous Estimates of Survival through Eight Years of Service using FY 1979 Cross-Sectional Data, by P. M. Lurie, September 1979.

5. Center for Naval Analyses Report, Estimating Four-Year Survival and Reenlistment Probabilities for Boiler Technicians and Machinist's Mates, by J. W. Fletcher, February 1979.

6. Center for Naval Analyses Report, The Effect of Delayed Entry, Recruit Quality and Training Guarantees for Two-Year Survival of Navy Enlisted Men, by R. F. Lockman, 1976.

7. Center for Naval Analyses Report CRC 382, First Term Survival and Reenlistment Chances for Navy Ratings and a Strategy for their Use, by J. S. Thomason, May 1979.

8. Nesbitt, K. W., The Development of Selection Standards for Three Navy Ratings which Vary in Level of Complexity, MS Thesis, Naval Postgraduate School, Monterey, June 1983.

9. Snyder, W. I. and Bergazzi, W. A., Enlistment Standards for Two Navy Ratings: Boiler Technicians (BT) and Machinist's Mates (MM), MS Thesis, Naval Postgraduate School, Monterey, June 1983.

10. McGarvey, W. E., Application of Covariance Structure Analysis to Enlistment Standards, paper presented at The Industrial Management Society and the Operations Research Society of America joint meeting , San Francisco, California, 15 May 1984.

11. Berenson, M. L. and Levine, D. M., Basic Business Statistics, 2nd ed., Prentice Hall, Inc., 1983.

12. SAS Institute Inc., SAS User's Guide: Statistics, 1982 ed., SAS Institute Inc., 1982.

13. Campbell, J. P., Psychometric Theory, in Dunnette, M. D., Handbook of Industrial and Organizational Psychology, Rand McNally, 1976.

14. Guilford, J. P. and Fruchter, B., Fundamental Statistics in Psychology and Education, 5th ed., McGraw-Hill, 1973.

15. Navy Personnel Research and Development Center Report TR 82-37, Prediction of Job Performance: Review of Military Studies, by R. Vineberg and J. N. Joyner, March 1982.

16. The Assessment Group Report R211, Billet Costs of Enlisted, Officer and Civilian Naval Personnel: FY 83, by O. L. Frankel and R. A. Butler, July 1983 (Updated March 1984).

17. Klecka, W. R., Discriminant Analysis, Sage Publications, Inc., 1980.

18. Nie, N. H. , and others, Statistical Package for the Social Sciences, 2nd ed., McGraw-Hill, 1975.

19. Wiggins, J. S., Personality and Prediction: Principles of Personality Assessment, Addison-Wesley, 1973.

BIBLIOGRAPHY

Center for Naval Analyses Report 82-1357, *Replacement Costs for Navy First Term Personnel by Rating*, by E. Balis and P. Clay-Mendez, September 1982.

General Research Corporation Report CR-197, *Development of Methods for Analysis of the Cost of Enlisted Attrition*, by D. F. Huck and R. D. Midlam, September 1977.

Sandel, C. D. and Gleason, M. F., *Enlistment Standards as Related to Performance in Aviation Antisubmarine Warfare Operator and Aviation Antisubmarine Warfare Technician Ratings*, MS Thesis, Naval Postgraduate School, Monterey, September 1983.

SAS Institute Inc., *SAS User's Guide*, 1982 ed., SAS Institute Inc., 1982.

The Assessment Group Report R207, *The Billet Cost Model System*, by R. A. Butler and C. L. Frankel, July 1983.

Whitmire, R. D. and Deitchman, C. G., *An Enlisted Performance Prediction Model for Aviation Structural Mechanics*, MS Thesis, Naval Postgraduate School, Monterey, August 1983.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3.	Department Chairman, Code 54 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
4.	Dr. Richard S. Elster Office of the Assistant Secretary of the Navy Room 4E778 The Pentagon Washington, DC 20350	1
5.	Professor William E. McGarvey, Code 54Ms Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	2
6.	Professor Thomas G. Swenson, Code 54Z Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
7.	Professor George W. Thomas, Code 54Te Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
8.	MPTA Library, Code 36 Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
9.	LCDR Dwayne A. Oslund (OP-130F1) Office of the Chief of Naval Operations Navy Department Washington, DC 20350	2
10.	CAPT J. S. A. Clark, RAE Personnel Branch-Army Russell Offices Canberra, ACT 2600 Australia	2
11.	Defence Library Campbell Park Offices Canberra, ACT 2600 Australia	2
12.	LT Charles G. Deitchman (OP-111E) Office of the Chief of Naval Operations Navy Department Washington, DC 20350	1

13. Professor Ronald A. Weitzman, Code 54Wz 1
Department of Administrative Sciences
Naval Postgraduate School
Monterey, California 93943

END

FILMED

5-85

DTIC